

THE
LONDON, EDINBURGH AND DUBLIN
PHILOSOPHICAL MAGAZINE
AND
JOURNAL OF SCIENCE

CONDUCTED BY

SIR LAWRENCE BRAGG, O.B.E., M.A., D.Sc., F.R.S.
PROF. G. P. THOMSON, M.A., D.Sc., F.R.S.

AND

ALLAN FERGUSON, M.A., D.Sc.

“Nec araneorum sane textus ideo melior quia ex se fila gignunt, nec
noster vilior quia ex allenis libamus ut apes.”

JUST. LIPS. *Polit. lib. i. cap. l. Not.*

VOL. XXXIV.—SEVENTH SERIES

JANUARY—DECEMBER 1943

LONDON:

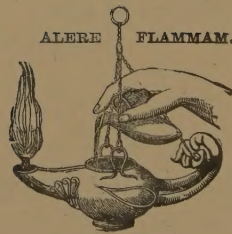
TAYLOR & FRANCIS, LTD., RED LION COURT, FLEET STREET, E.C. 4.

Agents: WM. DAWSON & SONS, LTD., 129 CANNON STREET, E.C. 4.

“Meditationis est perscrutari occulta; contemplationis est admirari perspicua Admiratio generat quaestionem, quaestio investigationem, investigatio inventionem.”—*Hugo de S. Victore.*

—— “Cur spirent venti, cur terra dehiscat,
Cur mare turgescat, pelago cur tantus amaror,
Cur caput obscura Phœbus ferrugine condât,
Quid toties diros cogat flagrare cometas,
Quid pariat nubes, veniant cur fulmina cœlo,
Quo micet igne Iris, superos quis conciat orbes
Tam vario motu.”

J. B. Pinelli ad Mazonium.



CONTENTS OF VOL. XXXIV. (Nos. 228-239)

(SEVENTH SERIES)

NUMBER 228—JANUARY

	Page
I. An Extension of Nomography. By C. W. HANSEL, Bedford School..	1
II. Modern Philosophy of Science and some Problems of Classical Physics. By G. BURNISTON BROWN, M.Sc., Ph.D.....	26
III. On Some Infinite Sets of Numbers. By LUDWIK SILBERSTEIN, Ph.D.	32
IV. Notes on the Reversion of a Series. By W. G. BICKLEY, D.Sc., Imperial College of Science and Technology, and J. C. MILLER, Ph.D., University of Liverpool	35
V. Formulæ Relating to Bessel Functions of Moderate or Large Argument and Order. By W. G. BICKLEY, D.Sc., Imperial College of Science and Technology	37
VI. Free and Forced Oscillations of Continuous Beams: Treatment by the Admittance Method. By Prof. W. J. DUNCAN, D.Sc., M.I.Mech.E., F.R.Ae.S., Wakefield Professor of Aeronautics at University College, Hull..	49
VII. The Separation of Electricity in Clouds. By J. ALAN CHALMERS, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham	63
VIII. On Solidified Water-films. By H. LÖWY, Dr.Phil.	67
IX. Note on a Passage in Stukeley's 'Memoirs of Sir Isaac Newton's Life.' By Prof. ALLAN FERGUSON, M.A., D.Sc.	71
X. Notices respecting New Books:—	
The Mathematical Tables Project	72

NUMBER 229—FEBRUARY

XI. Rational Electrodynamics.—I. The Limitations of Classical Electromagnetism. By E. A. MILNE, F.R.S.	73
XII. Rational Electrodynamics.—II. The Ideas of Kinematical Relativity. By E. A. MILNE, F.R.S.	82
XIII. Vapour-pressure and Electrodynamie Volume of Suspensions. By H. LÖWY, Dr.Phil.	101
XIV. The Boundary Couples in Thin Plates. By A. C. STEVENSON....	105
XV. The Torsion of a Fluted Column. By A. C. STEVENSON	115
XVI. The Repulsive Action between a Heated Solid Surface and the Adjacent Gas.—Part I. By H. SPENCER-GREGORY, Ph.D., A.R.C.S., D.I.C., Imperial College of Science and Technology	120

XVII. Further Notes on the Electron Density Distribution of the Upper Ionosphere. By OLOF E. H. RYDBECK, Sc.D., Chalmers Institute of Technology, Gothenburg, Sweden	130
XVIII. The Symmetrical Vortex Street in Sound-Sensitive Plane Jets. By P. SAVIC, Ph.D., and J. W. MURPHY	139

NUMBER 230—MARCH

XIX. The Hall Effect and some other Physical Constants of Alloys.—Part VII. The Aluminium-Silver Series of Alloys. By HANDEL POWELL, Ph.D., and Prof. E. J. EVANS, D.Sc., Physics Department, University College of Swansea.....	145
XX. The Approximate Theory of the Displacements produced in Encastred Beams due to the Elasticity of the Supports. By S. T. NEWING, B.Sc., A.R.C.S., D.I.C., The University, Bristol	162
XXI. Magnetic Moments and the Vibratory Electron. By IAN A. ROBERTSON, M.A., B.Sc., Research Student in the University of Glasgow	182
XXII. Rational Electrodynamics.—III. The Charge as Point Singularity. By Prof. E. A. MILNE, F.R.S.	197
XXIII. Fluorescence Notation. By JACK DE MENT.....	212
XXIV. Contact Potentials. By J. ALAN CHALMERS and W. HUME-ROTHERY.....	213
XXV. Notices respecting New Books :—	
LORD RAYLEIGH's Life of Sir J. J. Thomson	214

NUMBER 231—APRIL

XXVI. The Determination of Velocity of Sound by the Employment of Closed Resonators and the Hot-wire Microphone. By W. S. TUCKER, D.Sc., A.R.C.Sc.....	217
XXVII. Rational Electrodynamics.—IV. The "Radius" of a Point Charge. By E. A. MILNE, F.R.S.	235
XXVIII. Rational Electrodynamics.—V. The Neutron and Nuclear Dynamics. By E. A. MILNE, F.R.S.	246
XXIX. Voltage Fluctuations in Resistors. (Johnson Effect.) By N. R. CAMPBELL, Sc.D., F.Inst.P., and V. J. FRANCIS, B.Sc., A.R.C.S., F.Inst.P.....	259
XXX. Aristotle, Newton, Einstein. By E. T. WHITTAKER.....	266
XXXI. A Note on the Analysis of the Nuclear Binding Energies. By VLADIMIR VAND, D.Sc.....	280
XXXII. The Separation of Electricity in Clouds. By Sir GEORGE C. SIMPSON, F.R.S., Acting Superintendent, Kew Observatory.....	285
XXXIII. Note on the Paper "The Temperature Distribution around a Spherical Hole in an Infinite Conducting Medium," by Messrs. Pugh and Harris in the Phil. Mag. p. 661 (Sept. 1942). By H. S. CARSLAW ...	288

NUMBER 232—MAY

Page

XXXIV. On Sensory Energy, with special reference to Vision and Colour-Vision. By JOHN H. SHAXBY, D.Sc., F.Inst.P., Physiology Institute, University College, Cardiff	289
XXXV. The Evaluation of the Complex Roots of Algebraic Equations. By A. F. CORNOCK, M.Sc., and JOAN M. HUGHES. (Communication from the Staff of the Research Laboratories of The General Electric Company, Limited, England.)	314
XXXVI. The Absorption of Infra-red Radiation by Water-Vapour and Carbon Dioxide. By M. McCAIG, Ph.D., M.Sc., A.Inst.P., University of Manchester	321
XXXVII. The Reflexion of Electromagnetic Waves from a Parabolic Ionized Layer. By OLOF E. H. RYDBECK, Chalmers Institute of Technology, Gothenburg, Sweden	342
XXXVIII. The Single Potential Difference at a Cadmium Electrode. By J. ALAN CHALMERS, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham.....	349
XXXIX. The Stage Efficiency of an Impulse Generator. A Mathematical Note. By G. H. RAWCLIFFE (Robert Gordon's Technical College, Aberdeen, formerly of Liverpool University), M.A., A.M.I.E.E... ..	353
XL. Notices respecting New Books :—	
ARTHUR COX's Photographic Optics	360
Addendum. Corrections to Dr. G. F. C. Searle's paper in Phil. Mag. vol. xxxiii. Dec. 1942	360

NUMBER 233—JUNE

XLI. Standard Curves. By C. W. HANSEL, Bedford School.....	361
XLII. The Warming of Walls. By A. F. DUFTON, M.A., D.I.C.	376
XLIII. On Null Geodesics and Null-Corpuscles in the Theory of Relativity. By C. G. PENDSE, M.A., Ph.D.	377
XLIV. A Note on the Motion of a Mercury Index in a Capillary Tube. By H. GWYNEDD GREEN, M.A., University College, Nottingham	393
XLV. A Note on Entropy and Irreversible Processes. By Dr. W. EHRENBERG	396
XLVI. On the Iterative Solution of Linear Simultaneous Equations. By G. F. FREEMAN, M.Sc. (Eng.), A.M.I.E.E., A.M.I.Mech.E., West Ham Municipal College.....	409
XLVII. A Note on Stress Systems in Æolotropic Materials.—I. By A. E. GREEN	416
XLVIII. A Note on Stress Systems in Æolotropic Materials.—II. By A. E. GREEN	420
XLIX. Proofs of some Formulæ for the Hypergeometric Function and the E-function. By Prof. T. M. MACROBERT, Professor of Mathematics, University of Glasgow.....	422
L. Additional Note on the Trichromatic Theory of Colour Vision. (Referring to Phil. Mag., Aug. 1942, p. 559.) By Prof. W. PEDDIE	426

LI. Notices respecting New Books :—

Dr. J. L. HAUGHTON's Bibliography of the Literature relating to Constitutional Diagrams of Alloys	431
Mr. A. J. THOMSON's Table of the Coefficients of Everett's Central Difference Interpolation Formula	431

NUMBER 234—JULY

LII. The Properties of Fluids. By S. C. BRADFORD, D.Sc.	433
LIII. The Problem of the Keyboard Instrument. By LL. S. LLOYD ..	472
LIV. Three Types of Energy of Viscosity. By A. E. DUNSTAN, D.Sc., and A. H. NISSAN, Ph.D.	479
LV. Temperature, Pressure, and Specific Volume Changes of a Gas under Dissociation and Reassociation Conditions. By Professor W. J. WALKER, University of the Witwatersrand, Johannesburg, South Africa.....	486
LVI. The Representation of Intensity, Hue, Saturation, and the Magnitude of the Resultant Vector on the Trichromatic Theory. (With reference to recent Research on the Trichromatic Mechanism.) By Professor W. PEDDIE	488
LVII. The Fundamental Equation of Quantum Mechanics. By H. T. FLINT, Reader in Physics, University of London, King's College	496
LVIII. A Steady State Problem in Heat Conduction. By ARNOLD N. LOWAN.....	502

NUMBER 235—AUGUST

LIX. The Diode as a Rectifier and a Detector for Weak Signals. By J. AHARONI, King's College, London, at H. H. Wills' Laboratory, University of Bristol	505
LX. Reflexion and Transmission by Absorbing Dielectrics of Electromagnetic Waves in Hollow Tubes. By L. PINCHERLE, King's College, London. At the H. H. Wills Physics Laboratory, Royal Fort, Bristol....	521
LXI. Inversion Formulæ for the Laplace Transformation. By A. ERDÉLYI, Mathematical Institute of the University, Edinburgh	533
LXII. On the Capture of a Slow-moving Directed Electron in a Coulomb Field of Force. By W. R. MORGANS, D.Sc.	537
LXIII. Infinitesimal Bending of a Surface. By LUDWIK SILBERSTEIN, Ph.D.	549
LXIV. New Contributions to Interferometry. Part I.—New Non-Localized Interference Fringes. By S. TOLANSKY, Ph.D., D.I.C., Physics Dept., Manchester University. (Plates I.—III.)	555
LXV. Graphical Differentiation and Integration. By C. W. HANSEL, Bedford School	565
LXVI. Roots of $\sin z = z$. By A. P. HILLMAN and H. E. SALZER. (Mathematical Tables Project, National Bureau of Standards, U.S.A.)....	575
LXVII. Notices respecting New Books :—	
Messrs. M. RANDALL and L. YOUNG's Elementary Physical Chemistry	576
Messrs. R. A. FISHER and F. YATES's Statistical Tables for Biological, Agricultural and Medical Research	576

NUMBER 236—SEPTEMBER

	Page
LXVIII. Anomalous Viscosity shown in Oil Flow through Engine Bearings. By S. M. NEALE, D.Sc.....	577
LXIX. On the Dimensions of Physical Magnitudes (Third Paper: Electric and Magnetic Magnitudes). By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.	588
LXX. An Integral related to the Radiation Integrals. By E. O. POWELL	600
LXXI. The Duplex. By A. F. DUFTON, M.A., D.I.C.....	607
LXXII. Notes on Refraction at a Plane Interface and by a (Triangular) Prism. By ERIC J. IRONS, Ph.D., Queen Mary College, c/o King's College, Cambridge	608
LXXIII. On Contact Angles. By ERIC J. IRONS, Ph.D., Queen Mary College, c/o King's College, Cambridge.....	614
LXXIV. The Problem of the Keyboard Instrument.—II. By LL. S. LLOYD	624
LXXV. Preliminary Note on Condensation in the Form of Clouds and Dew. By J. W. ARCHBOLD, M.A., University College, London.....	632
LXXVI. Viscosity at the Boiling Point.—The Rheochor. By J. NEWTON FRIEND and WILLIAM D. HARGREAVES	643
LXXVII. Notices respecting New Books :—	
Prof. F. A. FISH's Fundamental Principles of Electric and Magnetic Circuits	650

NUMBER 237—OCTOBER

LXXVIII. Difficulties in Dirac's Representation Theory. By H. A. C. DOBBS	651
LXXIX. The Problem of the Keyboard Instrument.—III. By LL. S. LLOYD	674
LXXX. The Visual Spectrophotometry of Glasses, with Special Reference to Low Absorption Glasses. By T. H. WANG, M.Sc., D.I.C.....	684
LXXXI. The Significance of the Bakerian Lecture of 1843. By H. J. J. WINTER, M.Sc. (Lond.), A.Inst.P., University College of the South-West of England, Exeter.....	700
LXXXII. Note on the Interaction of Two Point-Charges. By E. A. MILNE, F.R.S.	712
LXXXIII. Distribution in Time of Spontaneous Fluctuation Voltage. By M. SURDIN, D.Sc.	716
LXXXIV. Notices respecting New Books.—	
Tables of Legendre Associated Functions. By ZAKI MURSI	722

NUMBER 238—NOVEMBER

	Page
LXXXV. Difficulties in Dirac's Representation Theory. By H. A. C. DOBBS	723
LXXXVI. Some Spectrophotometric Investigations on Iron Oxide-containing Soda-Lime-Silica Glasses.—Part I. The Influence of Concentration and Ferric-Ferrous Dissociation. By T. H. WANG, M.Sc., D.I.C., and Professor W. E. S. TURNER, F.R.S.....	744
LXXXVII. Some Boundary Problems of Two-Dimensional Elasticity. By A. C. STEVENSON.....	766
LXXXVIII. Notices respecting New Books :—	
R. A. ARCHIBALD'S Mathematical Tables and Aids to Computation.....	794

NUMBER 239—DECEMBER

LXXXIX. The Photographic Determination of Flame Temperatures in Closed-Vessel Explosions. By A. SMEETON LEAH, Ph.D., A.M.I.Mech.E. (Plate IV.)	795
XC. Electromagnet Polepieces and Condensers with Special Shapes. By N. DAVY, University College, Nottingham.....	803
XCI. Viscosities and Rheochors of Nitric Acid, Nitroparaffins and their Isomeric Nitrites. By J. NEWTON FRIEND and WILLIAM D. HARGREAVES	810
XCII. Unreleased Energy in Hydrocarbon Flame Gases. By W. T. DAVID, Sc.D., and J. MANN, B.Sc.....	816
XCIII. Some Properties of a Special Type of Electrical Pulse. By F. F. ROBERTS, B.Sc.(Eng.) and J. C. SIMMONDS, Ph.D.....	822
XCIV. The Principles of Carnot and Clausius. By Professor W. WILSON, F.R.S.	828
XCV. The Measurement of the Residual Parameters of Variable Air Condensers. By J. C. SIMMONDS.....	833
XCVI. Units and Dimensions. By HAROLD JEFFREYS.....	837
XCVII. On the Continuous Spectrum of Hydrogen-like Atoms (parabolic co-ordinates). By W. R. MORGANS, D.Sc.	842
Corrigenda to Dr. LUDWIK SILBERSTEIN'S Paper on the "Solution of the Equation $f'(x)=f(1/x)$," vol. xxx. (Sept. 1940) p. 186	850
Index	851

PLATES

- I.-III. Illustrative of Dr. S. TOLANSKY'S Paper on New Contributions to Interferometry. Part I.—New Non-Localized Interference Fringes.
- IV. Illustrative of Dr. SMEETON LEAH'S Paper on the Photographic Determination of Flame Temperatures in Closed-Vessel Explosions.

THE
LONDON, EDINBURGH & DUBLIN
PHILOSOPHICAL MAGAZINE
AND
JOURNAL OF SCIENCE

[SEVENTH SERIES—VOL. 34]

I. *An Extension of Nomography.*

By C. W. HANSEL, Bedford School *.

[Received August 10, 1942.]

1. *Introduction.*

NOMOGRAPHY has hitherto been very largely confined to the evaluation of a variable in an engineering formula. In this country, the subject has received tardy recognition and very little attention; indeed, many are unaware of its existence and some exhibit unreasoned hostility to its introduction into mathematical studies. It is unfortunate that nomographic methods have not been used more widely and more generally. There are probably a number of reasons why nomography has not become the preferred alternative to the slide-rule and mathematical tables as an instrument of calculation or as a method of reducing and exhibiting experimental data and tabular information. Two only need be mentioned. The generality of nomographic methods is not generally recognized and no facilities have been made available for their use. No attempt appears to have been made to employ nomographic methods systematically in the general solution of ordinary mathematical problems, although they are quicker and of more general application than those obtained with a slide-rule or squared paper. The user of a slide-rule or squared paper is not required to construct his slide-rule or to manufacture his squared

* Communicated by the Author.

paper, and the user of nomographic methods should be able to use nomographic paper ready made with scales printed on it ready for use.

The time and labour involved in constructing nomograms are considerable and are deterrents to the free use of nomographic methods. Nomographic calculation is not likely to find a place in elementary mathematical training until nomographic paper is available cheaply with ready-made scales printed on the paper. If paper of this kind becomes available at low cost it could with advantage be combined with squared paper and used even more extensively than the latter. It is not necessary to know the theory of nomographic methods before using them any more than it is necessary to know the theory of logarithms before these are used. Nomographic paper could be used to give a rapid check on every mathematical calculation, for the evaluation of $f(0)$, $f(1)$, etc., for curve plotting, for numerical differentiation and integration, numerical solution of differential and other equations, etc., etc. For most work of this kind only two general styles of nomographic paper need be used, and for elementary work only one.

The object of the present article is :—

1. To indicate the scope of nomographic methods in application to general mathematics.
2. To generalize nomographic methods, for example, to apply them to the representation and evaluation of imaginary quantities, to the numerical solution of differential equations, to the reduction of observations and the determination of laws.
3. To indicate by implication that early attention should be given to nomographic methods in any course of mathematical and scientific training.
4. To suggest the desirability of accurate nomographic paper being available for students undergoing a course of mathematical and scientific training.

The original intention of the author was to indicate how the accuracy of nomographic calculation could be substantially increased and to include detailed and accurately drawn nomograms. The high cost of publication renders this prohibitive. The diagrams throughout the paper are skeleton diagrams for purposes of explanation. They would be quite inadequate for purposes of nomographic computation.

2. Notation.

A NOMOGRAM is a system of lines, straight or curved and usually figured, used for obtaining numerical values of a quantity for any specified values of other quantities on which its value depends.

If a , b , --- x , y , --- are values of constant or variable quantities, a nomogram is for the purpose of evaluating $f(a, b, - - x, y, - -)$ for all values of a , b , --- x , y , ---.

A line, straight or curved, may be provided at intervals with graduation marks, some of which have a figure or number attached. The figured line is a SCALE and the line is the AXIS of the scale.

For simple nomographic work, figured lines may be numbered or lettered in pairs. The number or letter is the AXIS NUMBER or AXIS LETTER.

A number or letter enclosed by square brackets denotes a scale, the number or letter is in this case a SCALE NUMBER or SCALE LETTER.

For example: 1, 2, A, B, a , b , etc., may be used for denoting axes, [1], [2], [A], [B], [a], [b], etc., denote scales.

If scales are marked on sectional paper, and, as a rule, this is desirable, the intersections of the rulings with the scales may provide the graduation marks of a GENERAL SCALE, which may function nomographically with any or all of the scales marked on the paper. The figuring for a general scale may be marked along its axis or MARGINAL FIGURING may be used, the figures being usually placed to the left and above the graduations for FORWARD READING and to the right and below the graduation marks for BACKWARD READING.

A general scale may be numbered or lettered [0].

A scale used to graduate other scales is a BASE SCALE and may be denoted by a Roman numeral, for example: [I], [II], etc.

The graduation marks of a LATERAL SCALE meet the axis but do not pass through it.

The graduation marks of an AXIAL SCALE intersect the axis and extend to an equal distance on each side of it.

The order of marking scales may be indicated by ordinal subnumerals attached to the square bracket enclosing the scale number or scale letter. For example: $[3]_2$ means that scale number 3 is the second scale to be marked.

If there are two lateral scales, one to the left of the axis and the other to the right, these may be numbered 1 and 2, or 3 and 4, etc., or lettered A and A', B and B', etc. In the first case the axis of the scale would be lettered and in the second case when the scales are lettered the axis would be numbered if desired.

A UNIFORM SCALE has its consecutive graduations equidistant.

The POSITIVE DIRECTION of figuring a scale is the same as for Cartesian axes, *i.e.*, to the right or upwards from the zero.

A REVERSED SCALE has negative figuring, *i.e.*, the figures increase to the left or downwards from the zero.

If a reading a (figured or not) occurs on a scale l at a point A, this point or reading is denoted by $a[l]$, which symbolises the form of words "a point on scale l where the reading is a ," which may be abbreviated to " a on l ."

A normal to a scale or scales passes through OPPOSITE or CORRESPONDING READINGS or POINTS.

This is symbolised by " $a[l] \equiv b[m]$ ", which reads " a on l is opposite to b on m ."

If $a_1[l] \equiv b_1[m]$, and $a_2[l] \equiv b_2[m]$, then $\Delta a = (a_1 - a_2)$ and $\Delta b = (b_1 - b_2)$.

Δa and Δb are positive for a positively figured scale and negative for a negatively figured scale.

The figuring of a scale may be above, denoted by a , or below, b , the graduation lines, or it may be opposite the graduation line, o . The figures may be large black B, large red R, intermediate size black or red B' or R', or small black or red b or r .

The graduation marks may be long black or red B or R, intermediate length black or red B' or R', or short black or red b or r .

To describe the figuring and graduation of a scale, the figuring symbol is given first and the graduation symbol follows. For example, $aR'R$ denotes red figures of intermediate size above long red graduation lines, brb indicates small red figures below short black graduation lines.

The JOIN of two points $a[l]$ and $b[m]$ is denoted by $a[l] : b[m]$. If this join passes through the point $c[n]$, the join being produced if necessary, the ALIGNMENT of $c[n]$ with $a[l]$ and $b[m]$ is denoted by $a[l] : b[m] \rightarrow c[n]$. Thus the horizontal arrow with a double barb symbolises alignment or collinearity.

The horizontal arrow has a further significance. The arrow-head points towards the result of an arithmetical operation on the quantities from which the arrow is directed. In the example given, c is the numerical result of an arithmetical operation on a and b .

TRANSFERENCE of a reading on one scale $a[l]$ to another scale, where the equal reading is $a[m]$, is denoted by $a[l] \nearrow a[m]$.

A SECOR is an ungraduated and unfigured straight line which intersects one or more scales or REFERENCE LINES. A secor is a variable straight line and it may be given any desired translation and/or rotation. A secor is used for cross-alignment. If a secor intersects $[l]$ and $[m]$ at $a[l]$ and $b[m]$, this is indicated by $a[l] \leftharpoonup b[m]$, or " $a[l]$ aligns with $b[m]$." If the secor in the same position passes through $c[n]$, the mere collinearity of $a[l]$, $b[m]$, $c[n]$, is indicated by $a[m] \leftharpoonup b[m] \leftharpoonup c[n]$, but if c gives the result of an arithmetical operation of a on b , this is indicated by $a[l] : b[m] \rightarrow c[n]$.

A POINT on an unfigured REFERENCE LINE is denoted by $[R]$. For example, $a[l] \leftharpoonup b[m] \rightarrow [R] \leftarrow c[l] \leftharpoonup d[p]$ indicates that $a[l]$, $b[m]$, and $[R]$ are collinear and $c[l]$, $d[p]$, and $[R]$ are also collinear.

Sometimes it is necessary to rotate a secor about a PIVOT POINT. If the pivot point is on $[x]$ it is denoted by $Pa[x]$, but if the point is not on a numbered or a lettered scale it is written $P[R]$; in any case P denotes a pivot point.

If there are three parallel scales $[l]$, $[m]$, $[n]$, functionally associated so that $a[l] : b[m] \rightarrow f(a, b)[n]$, the perpendicular distance between $[l]$ and $[m]$ is $(\lambda + \mu)$, between $[l]$ and $[n]$ is λ , and between $[m]$ and $[n]$ is μ , and the scale distance ratio is λ/μ or μ/λ . If the scale zeros are collinear, and Δa , Δb , and Δc , are corresponding differences between pairs of opposite or corresponding readings:—

$$(\lambda + \mu) \frac{c}{\Delta c} = \frac{\lambda b}{\Delta b} + \frac{\mu a}{\Delta a} \text{ if } [n] \text{ is between } [l] \text{ and } [m],$$

and $(\lambda - \mu) \frac{c}{\Delta c} = \frac{\lambda b}{\Delta b} - \frac{\mu a}{\Delta a}$ if $[n]$ is not between $[l]$ and $[m]$,

and is on the side of $[m]$ remote from $[l]$.

$[l]$ and $[m]$ are PRIMARY SCALES and $[n]$ is a SECONDARY SCALE.

If a quantity is too large or too small for representation by a scale figure or reading, the scale figure may usually be multiplied by an appropriate power of 10 or 10^n to obtain the desired value for nomographic computation. In this case there is a FIGURE FACTOR of 10^n , the point $a[l]$ becomes $10^n a[l]$ or a figure factor of 10^n may be applied to the result.

It is sometimes convenient to write $|x|$ for $\log x$, $||x|$ for $\log \log x$, $|x$ for illog or $\text{antilog } x$, and $||x$ for $\text{illlog } x$.

Dy , D^2y , etc. denote dy/dx , d^2y/dx^2 , etc.

$D^{-1}y$ denotes $\int y \cdot dx$, and A, B, C, D , are constants of integration.

\neq denotes "is not equal to."

\doteq denotes "is approximately equal to."

3. Equations and Laws.

THE DIFFERENTIAL EQUATION $Dy=A$ may be regarded as the general equation of all straight lines. The solution is $y=Ax+B$ or $ax+by=c$, where a, b, c, x, y , may have any values.

Fig. 1.1 illustrates a family of straight lines $y=mx+c$ or $ax+by=c$, where $m=-a/b=-3/2$, the value of c being definite for each line.

Fig. 1.2 shows the family of lines $ax+by=c$, plotted to parallel coordinates. In this case, a and b have the same values as in 1.1, i.e., $a=3, b=2$, but for each value of c there is an infinite number of straight lines intersecting on the c -axis.

Fig. 1.3 indicates a method of drawing a nomogram for $ax+by=c$ when a, b, c , are given. Draw parallel axes x and y at unit distance apart. Draw the c -axis parallel to the axes of x and y so that the scale distance ratio $\lambda/\mu=b/a=-1/m$. Figure the axes. Only positive values of the parameters are shown but the figure is easily extended to negative values.

Fig. 1.3 gives a general numerical solution of the differential equation $Dy+m=0$ or $Dy=b/a$.

Fig. 1.4 indicates a method of finding the constants for a linear relation between tabular or observed values of x and y . (In 1.4, $x=C$ and $y=r$.) In this case draw the x and y scales unit distance apart and figure them; x negatively on the right, and y positively on the left.

Since $p[x]:q[y] \rightarrow r[c]$, the position of the c scale may be found by cross-alignment of corresponding values of x and y .

$\frac{c}{\Delta c} = \frac{\mu x}{\Delta x} + \frac{\lambda y}{\Delta y}$ gives the value of c corresponding to any pair of values

or opposite values of x and y .

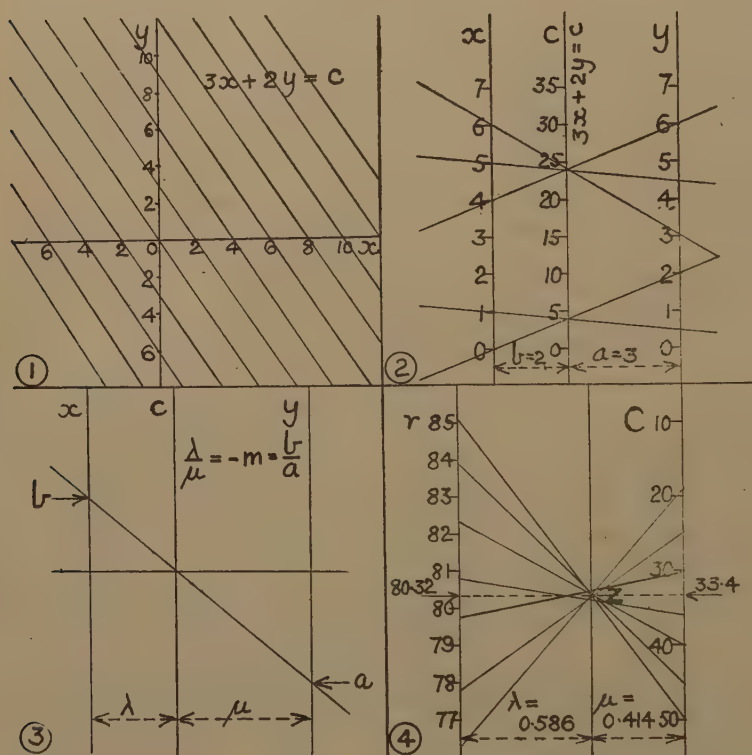
This type of nomogram furnishes a good method of testing whether a linear law satisfies observational data, and it provides an alternative and much less laborious method than the method of least squares for finding

the best values of the constants a , b , c . The method can also be used to eliminate the standard criteria of rejection by substituting ordinary visual examination of the intersecting lines on the nomogram.

Fig. 1.4 illustrates the nomographic method of finding the constants in a linear relation. The data are taken from 'Graphical and Mechanical Computation' by Joseph Lipka (1921).

The nomogram for working this problem (fig. 1.4 is a reduced similar nomogram) was drawn on Messrs. Pye's squared paper, ruled surface

Fig. 1.



50 cm. by 40 cm., the r and C scales being 40 cm. long and 50 cm. apart. For an exact linear law the secors should be concurrent. The diagram shows that some of the intersections are unsatisfactory, and these were rejected. The following calculation is based on guessing the point Z for it to lie at the centre of gravity of the six intersections lying closest together. It seems desirable to reject points whose deviation from the mean is excessive but, according to the method here adopted, rejection has been based on visual criticism of the points and not according to any standard criterion of rejection. In this way the point Z was found, giving:—

$$\lambda = 0.586, \quad \mu = 0.414, \quad r = 80.32, \quad C = 33.4.$$

$$\frac{z}{Az} = \frac{\mu r}{Ar} + \frac{\lambda C}{AC} = \frac{0.414r}{1} - \frac{0.586C}{5} = 0.414 \times 80.32 - 0.1172 \times 33.4 = 29.34,$$

$$0.414 r = 0.1172 C + 29.34, \quad r = 0.2831 C + 70.87.$$

The method of least squares applied to all the observations gives $r = 0.288 C + 70.76$, and applied to the four observations chosen gives $r = 0.288 C + 70.72$, the deviations in this case being $+0.07, -0.04, +0.09, -0.04$.

	C	r calculated	r observed	-deviation
Least squares	19.1	76.26	76.30	+0.04
formula	25.0	77.96	77.80	-0.16
$r = 0.288 C + 70.76$	30.1	79.43	79.75	+0.32
(Lipka)	36.0	81.13	80.80	-0.32
	40.0	82.28	82.35	+0.07
	45.1	83.76	83.90	+0.14
	50.0	85.16	85.10	-0.06
Nomogram	19.1	76.28	76.30	+0.02
formula	25.0	77.94	77.80	-0.14
$r = 0.283 C + 70.87$	30.1	rejected.		
	36.0	rejected.		
	40.0	82.19	82.35	+0.16
	45.1	rejected.		
	50.0	85.02	85.10	+0.08

If the method of least squares is preferred to the alternative method here given, the laborious computations involved in the application of the method of least squares are best performed with an arithmometer or may also be carried out nomographically by using Barlow's tables and logarithmic tables of considerable accuracy. Very few sets of experimental data are sufficiently numerous or sufficiently accurate to justify the application of the method of least squares. Moreover, the errors may be of such a nature as to make the method inadmissible. The treatment of observations by the method described in this paper enables a visual examination of the errors of observation to be made so that unsatisfactory data may be rejected.

Dy = Ax + B.

The primitive $y = ax^2 + bx + c$ represents, in Cartesian co-ordinates, all parabolas having their axes parallel to the axis of y .

Fig. 2.1 illustrates the Dy, x nomogram for $y = x^2 + 2x + 3 = ax^2 + bx + c$. $Dy = 2ax + b$. Hence, $Dy = b$ when $x = 0$.

Fig. 2.1 shows $Dy : x \rightarrow z$

$$(\lambda + \mu) \frac{z}{Az} = \frac{\mu \cdot Dy}{A(Dy)} + \frac{\lambda \cdot x}{Ax}.$$

If $(\lambda + \mu) = 1$ and Δz is chosen $= -\Delta(Dy) \cdot \Delta x$,

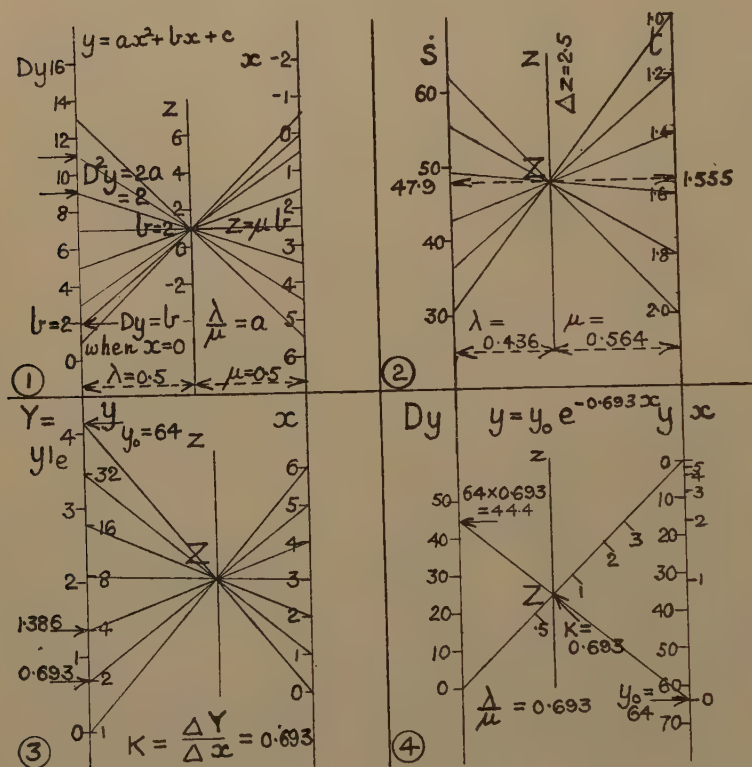
then $-z = \mu \cdot Dy \cdot \Delta x + \lambda \cdot x \cdot \Delta(Dy)$.

When $x=0$; $Dy=b$, and if in addition $\Delta x=-1$, then $z=\mu b$, which is true for any value of x since z is constant.

Again, $-\frac{dz}{dx} = \mu \cdot \Delta x \cdot D^2y + \lambda \cdot \Delta(Dy) = 0$ (since z is constant).

$\therefore D^2y = -\frac{\lambda \cdot \Delta(Dy)}{\mu \cdot \Delta x} = 2a$; hence, $a = \frac{\lambda}{\mu}$ if $\Delta x = -1$ and $\Delta(Dy) = 2$.

Fig. 2.



x	-1	0	1	2	3	4	5	6
y	2	3	6	11	18	27	38	41
$\Delta y = Dy$	1	3	5	7	9	11	13	
$\Delta^2 y = 2a$	2	2	2	2	2	2	2	

Fig. 2.2 illustrates the method of finding a quadratic law from experimental data.

Consider $y = y_0 \cdot e^{-\kappa x}$ or $\log y + \kappa x = \log y_0$ or $Y + \kappa x = Y_0$.

$\Delta Y = -\kappa \cdot \Delta x$ and $Y = Y_0$ when $x = 0$.

Fig. 2.3 shows the nomogram $\log y : x \rightarrow z$ or $Y : x \rightarrow z$,

$$(\lambda + \mu) \frac{z}{\Delta z} = \frac{\mu Y}{\Delta Y} + \frac{\lambda x}{\Delta x}.$$

If $(\lambda + \mu) = 1$ and $\Delta z = \Delta Y \cdot \Delta x$; then $z = \mu \cdot Y \cdot \Delta x + \lambda \cdot x \cdot \Delta Y$,

and
$$\frac{dz}{dx} = \mu \cdot \Delta x \cdot \frac{dY}{dx} - \lambda \cdot \Delta Y = 0,$$

$$\therefore \frac{dY}{dx} = -\frac{\lambda}{\mu} \cdot \frac{\Delta Y}{\Delta x} = \kappa \cdot \frac{\lambda}{\mu}.$$

Hence, $\Delta Y = \kappa$ when $\Delta x = 1$, which determines κ .

Fig. 2.4 illustrates the nomogram $Dy : y \rightarrow z$, where $Dy + \kappa y = 0$,

$$(\lambda + \mu) \frac{z}{\Delta z} = \frac{\mu \cdot Dy}{\Delta(Dy)} + \frac{\lambda \cdot y}{\Delta y}.$$

If $(\lambda + \mu) = 1$ and $\Delta z = \Delta(Dy) \cdot \Delta y$;

then
$$z = \mu \cdot Dy \cdot \Delta y + \lambda \cdot y \cdot \Delta(Dy),$$

or
$$z = (\mu \cdot \Delta y) \cdot -\kappa y + \lambda \cdot \Delta(Dy) \cdot y,$$

$$= y[-\kappa \cdot \mu \cdot \Delta y + \lambda \cdot \Delta(Dy)].$$

Since $z = 0$,
$$\kappa = -\frac{\lambda}{\mu} \cdot \frac{\Delta(Dy)}{\Delta y}.$$

If, as may well be the case, $\Delta(Dy) = \Delta y$; $\kappa = \frac{\lambda}{\mu}$.

After the nomogram $\log y : x$ or $Y : x$ has been drawn, κ becomes known. The nomogram $Dy : y$ may now be drawn using this value of κ . Hence, using both nomograms, any two of the quantities x , y , Dy may be found for any specified value of the third.

Fig. 2.2 relates to measurements of distance s and time t of a sphere rolling and oscillating under the action of gravity down a concave groove of radius appropriate to a period of oscillation of 0.40 second. ('Experimental Dynamics'—Hansel.) The sphere was a steel ball heavy enough to trace its own wave-trace by the aid of pen copying paper. The following measurements were made on a wave-trace so obtained:—

t secs.	.5	.7	.9	1.1	1.3	1.5
s cm.	4.30	8.03	12.96	19.12	26.42	35.00
$\dot{s} = \Delta s / \Delta t$ cm./sec.				30.70	36.79	42.94 49.07
t secs.	1.7	1.9	2.1	2.3	2.5	
s cm.	44.82	55.90	68.19	81.74	96.59	
$\dot{s} = \Delta s / \Delta t$ cm./sec.		55.32	61.59			

From a nomogram of \dot{s} and t (fig. 2.2),

$$\begin{aligned} \mu &= 0.564, & \lambda &= 0.436, & \lambda + \mu &= 1, \\ \dot{s} &= 47.9, & t &= 1.555, & \Delta \dot{s} &= 10, & \Delta t &= -0.25. \end{aligned}$$

$$\begin{aligned}\frac{z}{\Delta z} &= \frac{\mu \dot{s}}{\Delta \dot{s}} + \frac{\lambda t}{\Delta t} \\ &= \frac{0.564 \times 47.9}{10} - \frac{0.436 \times 1.555}{0.25} \\ &= 0.564 \times 4.79 - 1.744 \times 1.555 \\ &= -0.01036.\end{aligned}$$

$$0.0564 \dot{s} = 1.744 t - 0.01036$$

$$\dot{s} = 30.9 t - 0.184$$

$$s = 15.45 t^2 - 0.184 t + A.$$

Since $\dot{s} = 30.7$ when $t = 1.00$, $A = 0$.

$$\therefore s = 15.45 t^2 - 0.184 t.$$

$$\underline{Dy \pm ky = 0.}$$

Nomogram fig. 2.3 is $\log_e y \leftarrow x$ for the following data :—

x	0	1	2	3	4	5	6
y	64	32	16	8	4	2	1
$\log_e y$	4.16	3.47	2.77	2.08	1.39	0.693	0.000

The point Z on the nomogram gives the data for finding the relation between $\log_e y$ and x . It is found to be

$$\log_e y = -0.693 \cdot x + 4.16$$

or $\log_e y = -0.693 \cdot x + \log_e 64$, or $y = 64e^{-0.693x}$.

y may be evaluated for any value of x using nomogram 1 or 2.

For two consecutive values of x or $\Delta x = 1$, $\Delta y = 0.693$. This is the value of k . Hence the nomogram $Dy \leftarrow x$ can be constructed (fig. 2.4). Nomograms 2.3 and 2.4 enable any two of the quantities x , y , Dy to be found for any specified value of the third.

Nomogram fig. 3.1 relates to the following experiment :—

100 grams of water contained in a calorimeter of water equivalent 5 grams cooled in surroundings kept at a constant temperature $t_0 = 16.8^\circ \text{C}$. The temperature of the water was $t^\circ \text{C}$. at time T minutes.

T min.	0	1	2	3	4	5	6	7
$t^\circ \text{C}$.	47.7	46.8	46.0	45.2	44.5	43.8	43.2	42.5
$t^\circ \text{C}$. (corrected)	47.56	46.77	46.01	45.27	44.54	43.84	43.15	42.48
T min.	8	9	10	11	12	13	14	
$t^\circ \text{C}$.	41.8	41.1	40.5	40.0	39.4	39.0	38.4	
$t^\circ \text{C}$. (corrected)	41.82	41.19	40.57	39.96	39.37	38.80	38.24	

The nomogram is $\log(t - 16.8) \leftarrow T$ or $\log \theta \leftarrow T$. The readings for $T = 0$, 1, 13, and 14, are rejected as inaccurate. Using

$$(\lambda + \mu)z / \Delta z = \mu Y / \Delta Y + \lambda T / \Delta T$$

the nomogram gives the relation :—

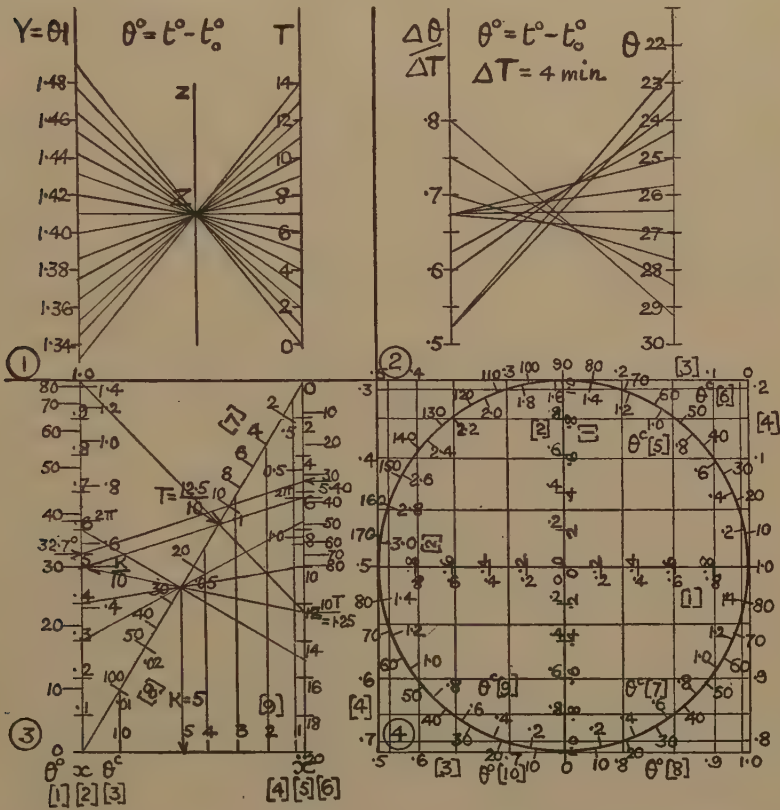
$$\log \theta = -0.01094 \cdot T + 1.486.$$

The observed and calculated values were as follows :—

T min.	0	1	2	3	4	5	6	7
$t^{\circ}\text{C. obs.}$	47.7	46.8	46.0	45.2	44.5	43.8	43.2	42.5
$t^{\circ}\text{C. calc.}$	47.4	46.66	45.92	45.19	44.48	43.80	43.12	42.46
δ_1	(0.3)	(0.14)	0.08	0.01	0.02	0.00	0.08	0.04
δ_2	0.14	0.03	0.01	0.07	0.04	0.04	0.05	0.02

T min.	8	9	10	11	12	13	14
$t^{\circ}\text{C. obs.}$	41.8	41.1	40.5	40.0	39.4	39.0	38.4
$t^{\circ}\text{C. calc.}$	41.83	41.21	40.60	40.01	39.43	38.87	38.27
δ_1	0.03	0.11	0.10	0.01	0.03	0.03	0.13
δ_2	0.02	0.09	0.07	0.04	0.03	0.20	0.16

Fig. 3.



The formula $\log \theta = -0.0112 \cdot T + 1.488$ may be applied to all the observations. δ_2 gives the deviations from this formula.

Nomogram fig. 3.2 shows $(-\Delta\theta/\Delta T)_{\Delta T=4} = -\theta$.

Evidently the data, though quite satisfactory, are not of sufficient accuracy to enable even a first difference to be used. Drawing tangents to the cooling curve to find the rate of cooling would give really bad results. Examiners sometimes set quite impossible questions relating to cooling observations.

The temperature observations are as accurate as it is possible to obtain them with a thermometer reading to 0.1° , the nomogram indicates what can be seen especially with a magnifier, the sticking of the mercury thread in the thermometer.

Errors of Observation.

It is not the purpose of this article to deal with the theory of errors but to indicate an alternative method to the method of least squares. When only a few observations of two quantities are available, which are subject to errors of various kinds, the method of least squares may not be the best method of finding the law of relation between the quantities or of finding the probable error of the result.

The importance of rejecting certain of the more inaccurate readings increases as the number of observations becomes smaller. Unless measurements are sufficiently accurate and numerous and of such a nature as to justify the application of the method of least squares and the standard criteria of rejection, these methods may not be so satisfactory as those based on experienced critical judgment. The nomographic method provides a visual analysis of the nature of the observation, and any relation between the observed quantities deduced nomographically depends on the judgment of selection of the weighted point Z.

4. Simple Harmonic Motion.

Cartesian and parallel co-ordinates are seldom used together. Some problems are readily solved by combining these two graphical methods.

NOMOGRAM 3.3 relates to the differential equation $\ddot{x} = -\kappa^2 x$.

$$x[2] : \ddot{x}[5] \rightarrow \kappa^2[7] \equiv \kappa[9].$$

$$\theta^0[1] \equiv \theta^c[3] \equiv \sin \theta[2] \text{ or } \cos (90 - \theta)[2].$$

$$\theta^0[6] \equiv \theta^c[4] \equiv \sin \theta[5] \text{ or } \cos (90 - \theta)[5].$$

If any two of the three quantities x , \ddot{x} , κ are given, the third may be found using 3.3.

$x = a \cos \theta = a \cos \kappa t$, where a is the amplitude of the motion.

If two of the quantities x , a , θ or θ , κ , t be known, the third may be found from 3.3.

The periodic time $T = \frac{2\pi}{\kappa}$ and the number of cycles per second $f = \frac{\kappa}{2\pi}$.

Example. $\ddot{x} + 25x = 0$.

$$\kappa = 5. \quad x[2] : \ddot{x}[5] \rightarrow 25[7] \equiv 5[9].$$

If $a = 0.1$, find x when $t = 0.2$.

$$0.2[2] : 50[7] \rightarrow 10[5] \text{ or } 10\theta^c[5]; \quad \theta^c = 1^\circ.$$

$$1^c[3] \equiv 57.3^0[1]; \quad 90 - 57.3 = 32.7.$$

$$32.7^0[1] : 10[7] \rightarrow 100x[5] \text{ or } 5.40[5]. \quad x = 0.0540.$$

Evidently if κ , a , and x are given, θ and t may be found, using 3.3.

Again,

$$\dot{x} = -a \sin \kappa t. \quad \text{Hence :}$$

$$a[7] : \theta^0[1] \text{ or } \theta^c[3] \rightarrow \dot{x}[5].$$

Also,

$$2\pi[5] : \kappa[2] \rightarrow T[7] \text{ and } f[8].$$

The circle diagram 3.4 relates to a particle having a periodic time $T=1$ second and amplitude $a=1.00$ units. The particle may move with uniform velocity v round the circle or it may move with simple harmonic motion along any diameter of the circle. For such a particle

$$x = \cos \theta = \cos 2\pi t \text{ or } y = \sin \theta = \sin 2\pi t.$$

The upper half of the diagram is figured for a particle having maximum displacement and zero velocity at time zero.

$$\theta^c[5] \equiv \theta^0[6] \equiv t[3] \equiv x[1] \equiv y[2] \equiv \cos \theta[1] \equiv \sin \theta[2].$$

The harmonically ruled paper enables corresponding values of these quantities to be read off without cross-alignment.

It is easy to adapt the diagram for the evaluation of the unknown quantities if $a \neq 1$ and $\kappa \neq 2\pi$, especially if 3.3 is used to give the product κt and the quotient x/a .

Example. Find x , \dot{x} , \ddot{x} , when $t=0.2$ if $\kappa=5$ and $a=0.1$.

$\theta^c = \kappa t$ is found from 3.3. $\kappa t = 1.0^c$ or 57.3^0 .

The circle diagram gives $\cos \theta = \frac{x}{a} = 0.540$ or $x = 0.0540$.

Also, $\sin \theta = y/a = 0.842$ or $\dot{x} = -0.0842$.

Nomogram 3.3 gives $\ddot{x} = -\kappa^2 x = -25 \times 0.0540 = -21.1$.

5. Numerical Differentiation and Integration.

NOMOGRAM 1 enables numerical values of various functions and their derivatives to be obtained.

This nomogram consists of eight logarithmic scales, [1], [2], - - - [8], two oblique and two horizontal scales [9], [10], [11] and [12], and two uniform axial scales [13] and [14]. The figuring and graduation of scales [1] to [12] is *oBB*, for [13] *bbb*, and for [14] *bbb*.

[1] is a positively figured logarithmic scale of range 1 to 100. [3], [5] and [7] are negatively figured logarithmic scales of range 1 to 1,000,000, 1 to 10,000, and 1 to 100 respectively.

[2], [4], [6], [8], are the reciprocal scales of [1], [3], [5], [7], respectively. For example,

$$x^n[5] \equiv x^{-n}[6]; \quad x^{-n}[5] \equiv x^n[6].$$

[9], [10], [11], [12], may be graduated by cross-alignment or, more accurately, by marking off the calculated distance of each graduation mark from one end of the scale.

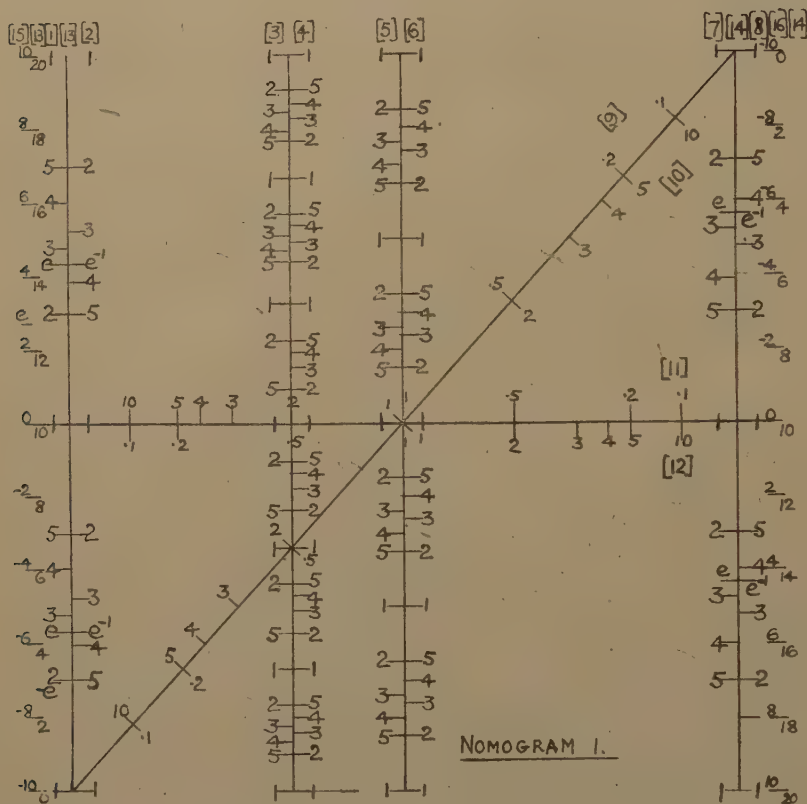
Let z = distance between the intersection of [1] and [9 or 11] and the intersection of [5] and [9 or 11], x = distance of graduation " a ."

Then, $x = \frac{2z}{a+2}$ for [9], and $= \frac{z}{a+1}$ for [11].

Thus the graduation figure 8[9] is at a distance 0.2 of the scale length (2) along the scale, measured from its point of intersection with [1].

Similarly, the distance of $a[10 \text{ or } 12]$ measured from [1] along [10] or [12] is $\frac{2az}{1+2a}$ for [10], or $\frac{az}{1+a}$ for [12]. Hence $-2[10]$ is at a distance $4z/3$ from the intersection of [1] and [10].

Nomogram 1.



$$y = f(x) = ax^n.$$

Nomogram 1 enables numerical values of ax^n to be found for all real values of x , a , and n .

For example:—

$$x[1] : (n+1)[9] \rightarrow x^n[5] \text{ or } x^{n-1}[3] \text{ or } x^{n+1}[7] \text{ or } x^{-n}[6] \\ \text{or } x^{1-n}[4] \text{ or } x^{-(1+n)}[8].$$

$$a[1] : x^n[5] \rightarrow ax^n[7] : b[5] \rightarrow ax^n/b[1].$$

$$n[1] : x^{n-1}[3] \rightarrow nx^{n-1}[5] : (n-1)[1] \rightarrow n(n-1)x^{n-1}[7].$$

$$nx^{n-1}[5] : a[1] \rightarrow anx^{n-1}[7] \text{ and } anx^{n-2}[5].$$

$$anx^{n-2}[5] : (n-1)[1] \rightarrow an(n-1)x^{n-2}[7] \text{ and } an(n-1)x^{n-3}[5].$$

$f(x)$ and its successive derivatives, direct or inverse, may be found rapidly by cross-alignment for all values of a , x , and n .

The evaluation of a definite integral involves two evaluations of the integral for two different values of x .

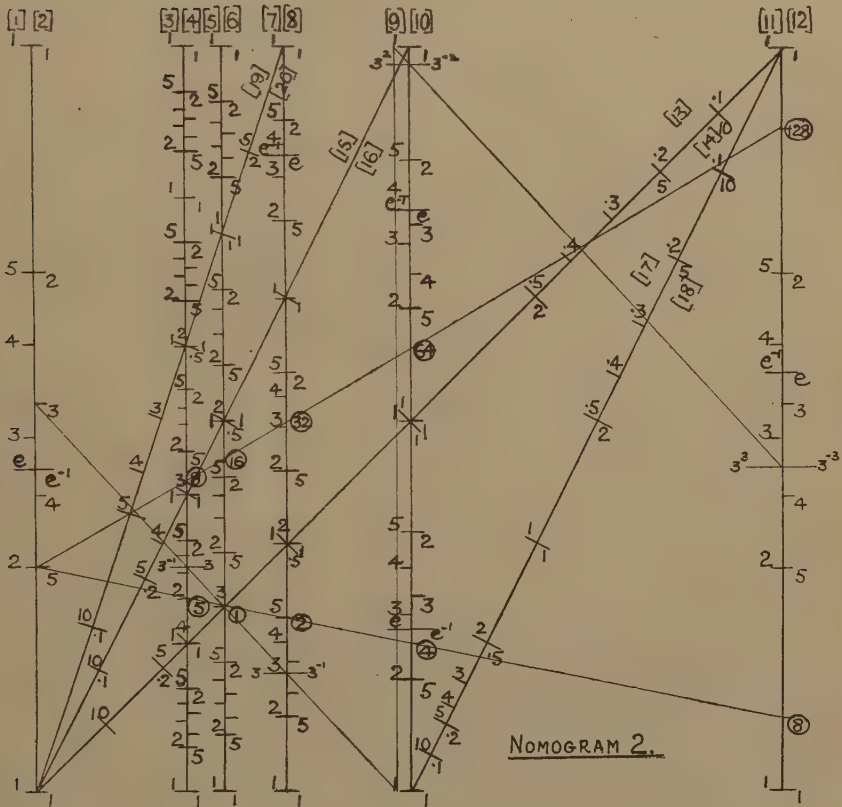
The value of x^n , x^{n-1} , x^{n+1} , may be obtained on any desired scale by adjustment of the value of n .

For example:— $x[1] : (n+1)[9] \rightarrow x^n[5]$ or $x^{-n}[6]$.

$x[1] : (n+2)[9] \rightarrow x^n[3]$ or $x^{-n}[4]$.

$x[1] : (n)[9] \rightarrow x^n[7]$ or $x^{-n}[8]$.

Nomogram 2.



If $y=f(x)=f_1(x)+f_2(x) \dots$, where $f_1(x)$, $f_2(x)$, etc., are each of the form Σax^n , y may be found by arithmetical addition of the values of $f_1(x)$, $f_2(x)$, etc., found nomographically. Hence the data for plotting $f(x) : x$, and for finding an approximate root of the equation $f(x)=0$, may be obtained from the nomogram. The method is limited to real values of a , x , n , but otherwise it is unrestricted.

Nomogram 2 indicates an arrangement of scales which give simultaneous

numerical values of various powers of x by a single cross-alignment.
For example :—

$$2[1] : 3[13] \rightarrow 2^0 \text{ or } 1[6] ; 2[8] ; 2^2[10] ; 2^3[12] ; 2^{-1}[4].$$

$$2[1] : 5[19] \rightarrow 2^5[8] ; 2^6[10] ; 2^8[12] ; 2^4[6] ; 2^3[4] ; 2^{-5}[7] ; \\ 2^{-6}[9] ; 2^{-7}[11] ; 2^{-4}[5] ; \text{ and } 2^{-3}[3].$$

$$3^{-1}[1] \text{ or } 3[2] : 3[13] \rightarrow 3^{-1}[3] ; 3[4] ; 3^0 \text{ or } 1[5 \text{ and } 6] ; 3^2[9] ; \\ 3^{-2}[10] ; 3^3[11] ; 3^{-3}[12].$$

$$y=f(x)=ae^{bx}.$$

Nomogram 1 enables numerical values of ae^{bx} and its successive derivatives to be found for all real values of a , b , and x .

For example :—

$$e^x. \quad e[1] : x[9 \text{ or } 11] \rightarrow e^x[7] \text{ and } e^{-x}[8].$$

$$e[7] : x[10 \text{ or } 12] \rightarrow e^x[1] \text{ and } e^{-x}[2].$$

$$bx. \quad x[1] : b[8] \rightarrow bx[6] \text{ or } b[1] : x[8] \rightarrow bx[6].$$

$$x[13] : b[9] \rightarrow bx[14] \text{ or } x[14] : b[10] \rightarrow bx[13].$$

$$ae^{bx}. \quad bx[6 \text{ or } 13 \text{ or } 14] \nearrow bx[9] : e[1] \rightarrow e^{bx}[7] :$$

$$e^{bx}[7] : a[2] \rightarrow ae^{bx}[5] \text{ or } ae^{-bx}[6].$$

$$Dy. \quad ae^{bx}[5] \nearrow ae^{bx}[1] : b[8] \rightarrow bae^{bx}[6].$$

$$D^{-1}y. \quad ae^{bx}[5] : b[2] \rightarrow ae^{bx}/b[7].$$

$$D^nae^{px}. \quad p[1] : 3[9] \rightarrow p^2[6], p^3[8], \text{ and } p^4[7].$$

$$ae^{px}[1] : p^n[8] \rightarrow p^nae^{px}[6].$$

Nomogram 2 may be used to advantage when several successive derived functions are required.

$$y=f(x)=a^x.$$

If $y=a^x$, $x=\log_a y$ or $y|_a$. Thus logs to any base may be found.

For example :—

$$x[7] : e[1] \rightarrow \log_e x \text{ or } x|_e[9].$$

$$x[7] : 10[1] \rightarrow \log_{10} x \text{ or } x|_{10}[9].$$

Lologs of numbers may also be found. Thus :—

$$x|_{10}[9] \nearrow x|_{10}[7] : 10[1] \rightarrow x|_{10}[9].$$

a^x and its successive derivatives are easily obtained.

$$a[1] : x[9] \rightarrow a^x[7] ; e[1] : a[7] \rightarrow a|_e[9] \nearrow a|_e[1] ;$$

$$a^x[8] : a|_e[1] \rightarrow a^x. a|_e[6].$$

Other derivatives are obtained similarly.

6. Circular Functions.

NOMOGRAM 4.1 is for evaluations relating to :—

$$a \cdot \cos \theta + b \cdot \sin \theta = r \cdot \cos \phi = c,$$

where

$$r = \sqrt{(a^2 + b^2)}, \quad \phi = \theta - \alpha, \quad \alpha = \tan^{-1} b/a.$$

An appropriate figure factor enables the nomogram to be used for all values of a , b , and c .

[1], [3], [5] are uniform scales figured so that :—

$$a[1] : [b]5 \rightarrow (a+b)3.$$

[2], [4], [6] are "square" scales figured so that :—

$$a[2] \equiv a^2[1], r[4] \equiv (a^2 + b^2)[3], b[6] \equiv b^2[5].$$

Hence, $a[2] : b[6] \rightarrow r[4]$ or $\sqrt{(a^2 + b^2)}[4]$.

[7] and [8] are graduated so that $\theta^c[8] \equiv \cos \theta[7]$,
and $a[4] : \cos \theta[7] \rightarrow (a \cdot \cos \theta)[2]$.

[9] and [10] are graduated so that $\theta^0[10] \equiv \sin \theta[9]$,
and $b[4] : \sin \theta[9] \rightarrow (b \cdot \sin \theta)[6]$.

If x = distance between $\cos \theta[7]$ and $0[1]$, and z = distance between $0[1]$ and $0[3]$,

$$\frac{x}{z-x} = \frac{2a^2}{a^2 + b^2}, \therefore x = \frac{2a^2 z}{3a^2 + b^2} \text{ or } \frac{2z}{3 + \tan^2 \theta} \text{ or } x = \frac{2 \cos^2 \theta \cdot z}{2 \cos^2 \theta + 1},$$

where $\theta = \tan^{-1} b/a$.

Similarly, if x = distance between $\sin \theta[9]$ and $0[6]$, and z = distance between $0[6]$ and $0[3]$,

$$x = \frac{2 \sin^2 \theta \cdot z}{2 \sin^2 \theta + 1}, \text{ where } \theta = \tan^{-1} b/a.$$

To graduate [7], evaluate x for values of $\cos \theta = 0.1, 0.2$, etc.

To graduate [9], evaluate x for values of $\sin \theta = 0.1, 0.2$, etc.

To graduate [8], evaluate x for values of $\theta^c = 0.1, 0.2$, etc.

To graduate [10], evaluate x for values of $\theta^0 = 10^0, 20^0$, etc.

$$a \cdot \cos \theta + b \cdot \sin \theta = r \cdot \cos \phi = c.$$

$$r = \sqrt{(a^2 + b^2)}, \phi = \theta - \alpha, \alpha = \tan^{-1} b/a.$$

1. Given a, b, θ ; to find r, ϕ, α, c

$$a[2] : b[6] \rightarrow r[4].$$

$$a[4] : \theta^c[8] \text{ or } a[4] : \cos \theta[7] \rightarrow a \cdot \cos \theta[2] \nearrow a \cdot \cos \theta[1].$$

$$b[4] : \theta^0[10] \text{ or } b[4] : \sin \theta[9] \rightarrow b \cdot \sin \theta[6] \nearrow b \cdot \sin \theta[5].$$

$$a \cdot \cos \theta[1] : b \cdot \sin \theta[5] \rightarrow (a \cdot \cos \theta + b \cdot \sin \theta) \text{ or } c[3],$$

$$c \text{ or } r \cdot \cos \phi[3] \nearrow r \cdot \cos \phi[2] : -r[4] \rightarrow \cos \phi[7] \text{ and } \phi^c[8].$$

$$-r[4] : b[6] \rightarrow \sin \phi[9] \text{ and } \phi^0[10].$$

$$\alpha = \phi - \theta \text{ or } \phi[3] : \theta[1] \rightarrow \alpha[5].$$

2. Given r, ϕ, α ; to find a, b, θ , and c .

$$\theta = \phi - \alpha \text{ or } \phi[3] : \alpha[5] \rightarrow \theta[1].$$

$$-r[4] : \cos \phi[7] \rightarrow r \cdot \cos \phi \text{ or } c[2].$$

$$-r[4] : \alpha[8] \text{ or } -r[4] : \cos \alpha[7] \rightarrow a \text{ or } r \cdot \cos \alpha[2].$$

$$-r[4] : \alpha[10] \text{ or } -r[4] : \sin \alpha[9] \rightarrow b \text{ or } r \cdot \sin \alpha[6].$$

3. Given a, b, c ; to find θ .

$$a[2] : b[6] \rightarrow r[4].$$

$$c \text{ or } r \cdot \cos \phi[2] : -r[4] \rightarrow \cos \phi[7] \text{ or } \phi^c[8].$$

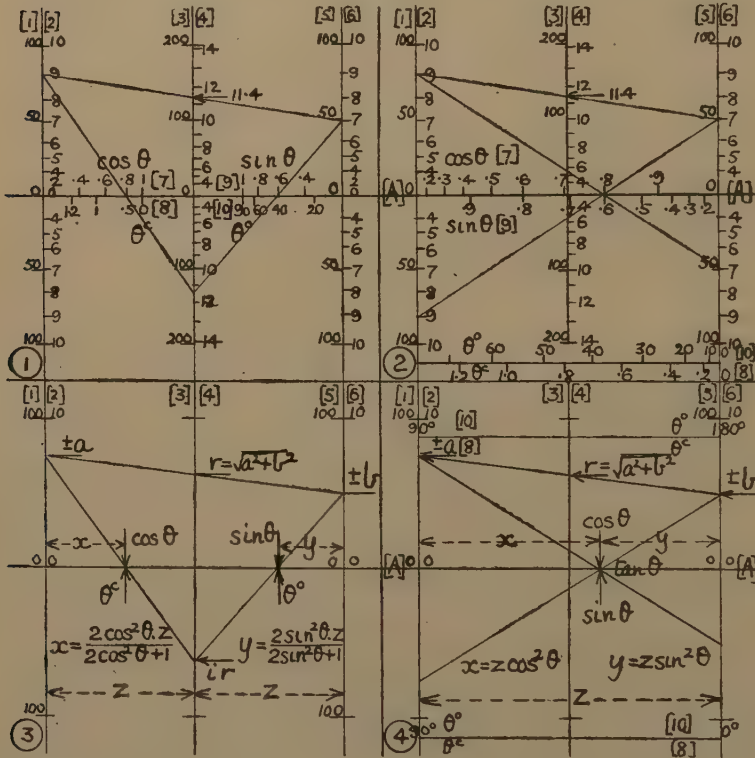
$$a \text{ or } r \cdot \cos \alpha[2] : -r[4] \rightarrow \cos \alpha[7] \text{ or } \alpha^c[8].$$

$$\theta = \phi - \alpha.$$

NOMOGRAM 4.2 is similar to Nomogram 4.1, but the scales for $\cos \theta$, $\sin \theta$, and θ extend from 0[1] to 0[5]. The scales along the horizontal axes are as follows:—

$$\cos \theta[7] \equiv \theta^c[8] \equiv \theta^o[10] \equiv \sin \theta[9].$$

Fig. 4.



Nomogram 4.2 gives corresponding values of θ^o , θ^c , $\cos \theta$, $\sin \theta$.

Also, $a[2] : b[6] \rightarrow \sqrt{a^2 + b^2}[4] \text{ or } \cos \theta[7] \text{ or } \sin \theta[9],$

where $\theta = \tan^{-1} b/a$.

NOMOGRAM 3 is useful to obtain products, powers, and reciprocals of circular functions. [4], [8], [9], [28] are positively figured logarithmic scales. [1], [5], [12], [29] are the corresponding reciprocal scales. Only a portion of [28] and [29] are shown.

$$\theta^c[2] \equiv \cos \theta[A] \equiv a[3].$$

$$\theta^0[9] \equiv \sin \theta[C] \equiv b[11].$$

$$a[3] : b[11] \rightarrow (a^2 - b^2)[7],$$

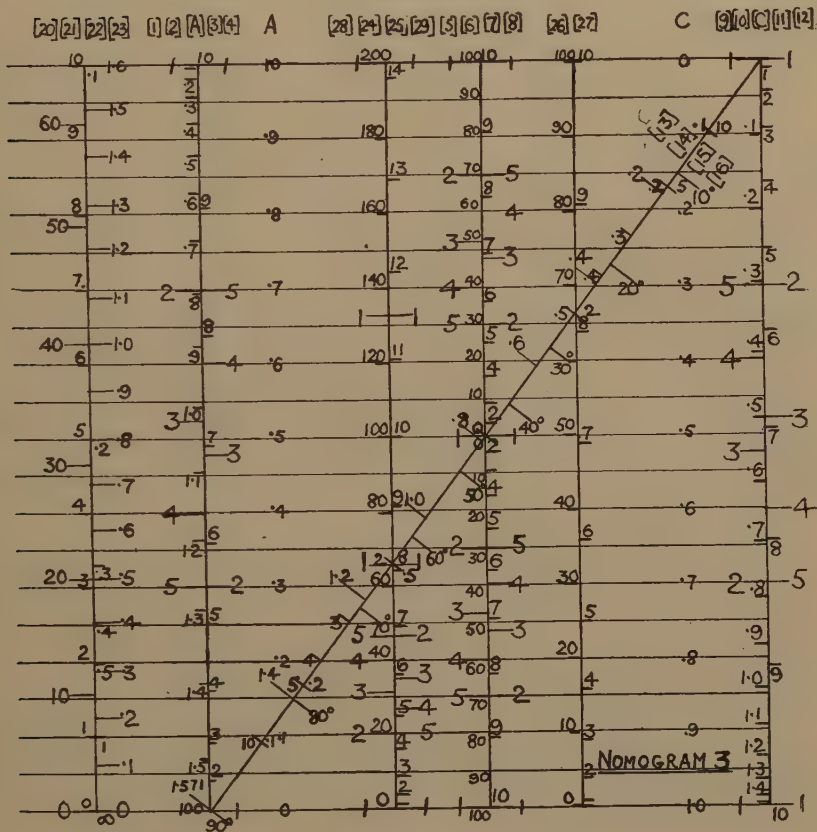
$$a[3] : b[27] \rightarrow r \text{ or } \sqrt{(a^2 + b^2)}[25],$$

$$x[4] : y[12] \rightarrow x/y[8].$$

Hence $\cos^2 \theta - \sin^2 \phi$, $\cos \theta / \sin \phi$, $(\cos \theta - \cos \phi)$, etc., may be evaluated. The oblique scales [13], [14], [15], [16], are as follows:—

$$\theta^0[13] \equiv \tan \theta[14] \equiv \cot \theta[15] \equiv \theta^0[16].$$

Nomogram 3.



[14] and [15] are segmentary scales, hence

$$a[A] : c[14] \rightarrow ac[C], \quad b[C] : c[15] \rightarrow bc[A].$$

$$\cos \theta[A] : 1[10] \rightarrow \sec \theta[14], \quad \sin \theta[C] : 1[2] \rightarrow \operatorname{cosec} \theta[15].$$

$$\text{Alternatively, } \cos \theta[4] \equiv \sec \theta[2].$$

Products of trigonometrical functions may be found by successive cross-alignments, a point being transferred to another axis if necessary.

$$x[4] : a[14] \rightarrow x^a[12] \text{ or } x^{-a}[9].$$

Hence, powers of trigonometrical functions may be found; for example, $\cos^n \theta$, $\tan^n \theta$, etc.

[20], [21], [22], [23] are for use in connexion with periodic functions. Quantities often occurring in calculations are, f =frequency, $p=2\pi f$, $q=\frac{f}{2\pi}$, and periodic time $T=1/f$.

[20], [21], [22], [23] are graduated as follows:—

$$p[20] \equiv f[21] \equiv T[22] \equiv q[23].$$

If $x=a \cdot \cos . pt$, x and its derivatives are easily evaluated with the aid of Nomogram 3. Appropriate figure factors may have to be used to give a suitable range of values.

[24] and [25] enable this nomogram to be used for evaluations relating to $(a \cdot \cos \theta + b \cdot \sin \theta)$, using the methods already given for Nomogram 4.2.

$$a[3] : b[11] \rightarrow (a^2 + b^2)[24] \text{ and } r \text{ or } \sqrt{(a^2 + b^2)}[25].$$

[3], [2], [29], [28], [8], [6], [12], [10] of nomogram 3 are the same as [1], [2], [3], [4], [5], [6], [7], [8] of nomogram 1.

Hence, the calculations which can be performed on nomograms 1 and 4.2 can all be carried out with nomogram 3.

Nomogram 3 is a general purpose nomogram, capable of carrying out all general types of calculation.

7. Complex Quantities.

NOMOGRAM 4.1 may be used for the representation of complex numbers. [1], [3], [5], are uniform scales figured $a^2[1] \equiv (a^2 + b^2)[3] \equiv b^2[5]$.

[2], [4], [6] are square scales and give the corresponding values of a , $r = \sqrt{(a^2 + b^2)}$, b , respectively.

$$\pm a[2] \equiv a^2[1], \pm r[4] \equiv (a^2 + b^2)[3], \pm b[6] \equiv b^2[5].$$

$$\pm a[2] : \pm b[6] \rightarrow r^2[3] \text{ or } \sqrt{(a^2 + b^2)}[4] \text{ or } (a^2 + b^2)[3].$$

Positive and negative values of a and b are coincident on the upper portions of [2] and [6].

Since the operator i^2 turns any vector a^2 through 180° ,

$$(i \cdot a)[2] \equiv (ia)^2[1] \text{ or } -a^2[1],$$

$$(i \cdot b)[6] \equiv (ib)^2[5] \text{ or } -b^2[5].$$

The complex number $(a + i \cdot b)$ is represented by the join of $a[2]$ to $(ib)[6]$ or $a[2] : (ib)[6]$.

$$a[2] : (ib)[6] \rightarrow \sqrt{(a^2 - b^2)}[4].$$

If $a[2] : b[6]$ is produced to meet the line joining the zero of the scales, the point of intersection may be figured $\tan \theta$ and θ^c , where $\theta^c = \tan^{-1} b/a$. If a and b are given, r and θ can be read on the nomogram, and conversely if r and θ are given, a and b may be obtained.

If x =distance of point $\tan \theta[7]$ from $0[1]$, and z =distance between $0[1]$ and $0[6]$,

$$\frac{x}{z-x} = \frac{a^2}{b^2} \text{ or } x = z \frac{a^2}{a^2 + b^2} = \frac{z}{1 + \tan^2 \theta}.$$

[7] may be graduated so that $a[2] : b[6] \rightarrow \tan \theta[7]$.

$\tan \theta[7] \equiv \theta^c[8]$ or $\theta^o[10]$.

Alternatively, [7] and [9] may be graduated as in nomogram 4.1 for $\cos \theta$ and $\sin \theta$ respectively, so that,

$a[2] : b[6] \rightarrow r$ or $\sqrt{(a^2 + b^2)}[4]$,

$(ir)[4] : a[2] \rightarrow \cos \theta[7]$ and $\theta^c[8]$,

$(ir)[4] : b[6] \rightarrow \sin \theta[9]$ and $\theta^o[10]$.

Then, $(a + i \cdot b) = r \cdot e^{i\theta} = r(\cos \theta + i \cdot \sin \theta)$. To express $r \cdot e^{i\theta}$ in the form $(a + i \cdot b)$ or $r(\cos \theta + i \cdot \sin \theta)$:—

$(ir)[4] : \theta^c[8] \rightarrow a[2]$ and $\cos \theta[7]$,

$(ir)[4] : \theta^o[10] \rightarrow b[6]$ and $\sin \theta[9]$.

NOMOGRAM 4.2 gives more open scales for $\cos \theta$ and $\sin \theta$.

$a[2] : b[6] \rightarrow \cos \theta[7]$, $\sin \theta[9] \equiv \theta[8 \text{ or } 10]$.

$\tan \theta[A] \equiv \theta^c[8] \equiv \theta^o[10]$.

Hence, if a and b are given, r and θ may be found.

If r and θ are given,

$\theta^c[8] \equiv \tan \theta[A] \equiv \cos \theta[7] \equiv \sin \theta[9]$.

Then, $a = r \cdot \cos \theta$, $b = r \cdot \sin \theta$.

A better method of dealing with this problem is given later.

Figs. 4.3 and 4.4 indicate how to mark [A], [7], [8], [9] and [10].

FIGURE 5.1.—Fig. 5.1 is a vector diagram of the complex numbers $[a, b]$, $[-a, b]$, $[-a, -b]$, $[a, -b]$. r has the same positive value for each of these vectors.

Fig. 5.3 indicates how to construct a nomogram for expressing a complex number $(a + i \cdot b)$ in terms of hyperbolic functions.

$[a, b] = (a + i \cdot b) = r(\cosh \cdot u + i \cdot \sinh \cdot u)$,

$r = \sqrt{(a^2 - b^2)}$, $u = \tanh^{-1} b/a$.

NOMOGRAM 3 is the most convenient for purposes of evaluation.

1. To express $(a + i \cdot b)$ in the form $r \cdot e^{i\theta}$.

$a[7] : b[11] \rightarrow r[3]$, $a[A] : b[C] \rightarrow \theta^c[13]$.

2. To express $r \cdot e^{i\theta}$ in the form $(a + i \cdot b)$.

$r[14] : \theta^c[2] \rightarrow a[C]$, $r[15] : \theta^o[10] \rightarrow b[A]$.

3. To evaluate $(a + i \cdot b)(c + i \cdot d)$.

Express $(a + i \cdot b)$ and $(c + i \cdot d)$ in the forms $r_1 \cdot e^{i\theta}$ and $r_2 \cdot e^{i\phi}$.

$r_1[4] : r_2[9] \rightarrow r_1 r_2[8]$.

$(a + i \cdot b)(c + i \cdot d) = r_1 r_2 \cdot e^{i\psi}$, $\psi = \theta + \phi$.

4. To evaluate $(a + i \cdot b)/(c + i \cdot d)$.

$r_1[4] : r_2[8] \rightarrow r_1/r_2[12]$.

$(a + i \cdot b)/(c + i \cdot d) = r_1/r_2 \cdot e^{i\psi}$, $\psi = \theta - \phi$.

5. To evaluate $(a+i \cdot b)^n$.

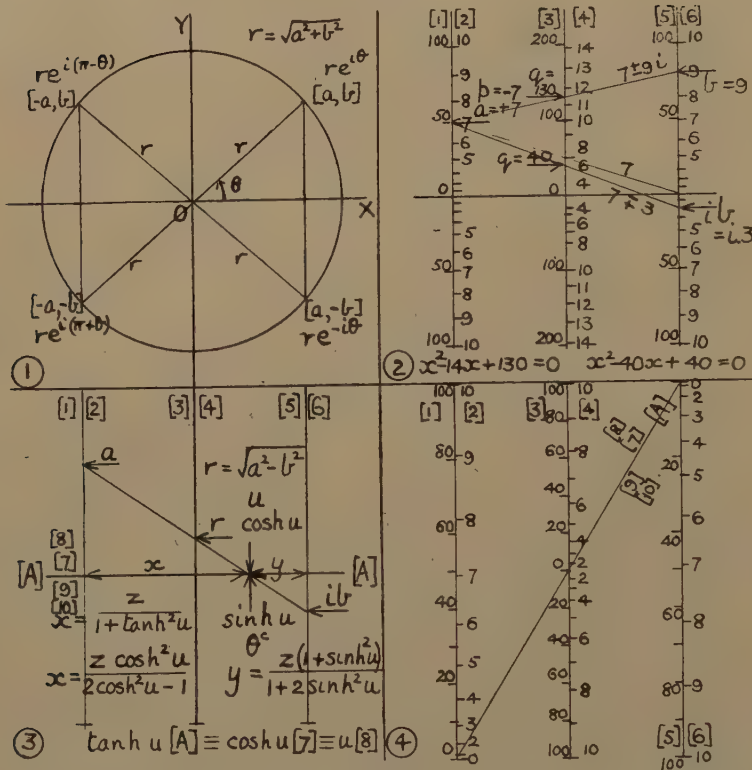
$$=r^n \cdot e^{in\theta}.$$

$$r[4] : n[14] \rightarrow r^n[12], \theta[4] : n[9] \rightarrow n\theta[8].$$

6. To evaluate $(a+i \cdot b)^{\frac{1}{n}} = r^{\frac{1}{n}} \cdot e^{i\frac{\theta}{n}}$.

$$r[4] : n[8] \rightarrow r/n[12], \theta[4] : n[8] \rightarrow \frac{\theta}{n}[12].$$

Fig. 5.



7. To evaluate $(a+i \cdot b)^{-n} = r^{-n} \cdot e^{-in\theta}$.

$$r^n[14] \equiv r^{-n}[15], \theta[4] : n[9] \rightarrow n \cdot \theta[8].$$

There are obvious alternative methods.

8. Quadratic Equations.

Consider a quadratic equation having real coefficients. Let the roots of the equation be $\alpha = (a+i \cdot b)$, $\beta = (a-i \cdot b)$.

The equation is $x^2 - 2ax + (a^2 + b^2) = 0$, the term in x^2 being positive.

If $\alpha = (-a+i \cdot b)$ and $\beta = (-a-i \cdot b)$, the equation is $x^2 + 2ax + (a^2 + b^2) = 0$.

If α and β are real, b is imaginary since a is real.

If the equation is written in the form $x^2 \pm 2px + q = 0$, $p = \mp a$ and $q = (a^2 + b^2)$.

p and q being real, a has the opposite sign to p .

NOMOGRAM 5.2 is for the solution of equations of the form $x^2 \pm 2px + q = 0$, the roots being $\alpha = (a + i \cdot b)$ and $\beta = (a - i \cdot b)$.

If the equation is not already in this form, a nomogram for division of numbers may be used to reduce the equation to the required form.

[1], [3], [5] are graduated uniformly.

$(-a)$ or $p[2] \equiv a^2[1]$, $b[6] \equiv b^2[5]$,

$\pm a[2] : b[6] \rightarrow (a^2 + b^2)$ or $q[3] \equiv \sqrt{(a^2 + b^2)}[4]$.

Positive and negative values of a and b are coincident on the upper half of [2] and [6]. The lower halves of the scales represent values of $(i \cdot a)[2]$ and $(i \cdot b)[6]$.

$(i \cdot a)[2] \equiv -a^2[1]$, $(i \cdot b)[6] \equiv -b^2[5]$

To solve a quadratic equation, it is reduced to the form $x^2 \pm 2px + q = 0$, the coefficient of x^2 being +1.

Then, $a = \mp p$, *i. e.*, a has opposite sign to p .

$\pm p[2] : q[3] \rightarrow b[6]$ and $\mp a[2]$, giving $\alpha = a + i \cdot b$, $\beta = a - i \cdot b$.

Example :—

$$x^2 - 14x + 40 = 0 \text{ or } x^2 - 2 \times 7x + 40 = 0.$$

$$a = 7, q = 40.$$

$$7[2] : 40[3] \rightarrow i \cdot 3[6]$$

$$\alpha = 7 - 3 = 4, \beta = 7 + 3 = 10.$$

The addition and subtraction of a and $\pm i \cdot (i \cdot b)$ may be carried out nomographically using [1], [3], and [5], if desired.

Nomogram 5.2 gives a general solution of all quadratic equations having real coefficients. If the values of p and q do not fall within the range of figuring of the scales, a figure factor of 10^n may be used on all scales except 4, for which the figure factor would be 10^{2n} , n being a positive or negative integer of appropriate value.

$D^2(y) + 2p \cdot D(y) + q \cdot y = 0$, p and q are constants.

Let $\alpha = (a + i \cdot b)$ and $\beta = (a - i \cdot b)$ be the roots of the quadratic equation $\lambda^2 + 2p\lambda + q = 0$.

If A and B are arbitrary constants whose values may be found from special conditions, the solution of the differential equation is :—

$y = A \cdot e^{\alpha x} + B \cdot e^{\beta x}$ if $\alpha \neq \beta$ and the roots are real. y may be evaluated for all values of a , b , α and β with the aid of nomogram 3.

$y = A \cdot e^{\alpha x} + B \cdot x \cdot e^{\beta x}$ if $\alpha = \beta$ and the roots are real. Nomogram 3 again gives a complete numerical solution.

$y = A \cdot e^{\alpha x} \cdot \sin(bx + B)$ if α and β are imaginary. In this case a general numerical solution may be obtained, using nomogram 3.

The values of A and B can, in general, be evaluated using nomogram 3. Hence, nomograms 5.2 and 3 give all numerical solutions of the equation.

9. Differential Operators.

$$D^n(y) = D^n(a \cdot \sin \cdot bx) = a \cdot b^n \cdot \sin \left(bx + \frac{2n\pi}{4} \right). \quad (\text{Nomogram 3.})$$

$$2n\pi[20] \equiv n[21], b[4] : x[9] \rightarrow bx[8].$$

It is easier to evaluate $\left(bx + \frac{2n\pi}{4} \right)$ arithmetically than nomographically, but the evaluation is possible with the aid of nomogram 3.

$$b[4] : n[14] \rightarrow b^n[12] \nearrow b^n[4],$$

$$b^n[4] : a[9] \rightarrow a \cdot b^n[8] \nearrow a \cdot b^n[4].$$

$$\left(bx + \frac{2n\pi}{4} \right)[9] : a \cdot b^n[4] \rightarrow a \cdot b^n \cdot \sin \left(bx + \frac{2n\pi}{4} \right)[8].$$

Similarly, $D^n(a \cdot \cos \cdot bx) = a \cdot b^n \cdot \cos \left(bx + \frac{2n\pi}{4} \right)$ may be evaluated for all values of a, b, n .

$$\text{Or, } D^n(y) = (i \cdot b)^n y = e^{\frac{in\pi}{2}} \cdot b^n \cdot y.$$

$$\text{If } y = a \cdot \sin \cdot bx, D^n y = a \cdot b^n \cdot \sin \left(bx + \frac{n\pi}{2} \right) \text{ or } a \cdot b^n \cdot \sin \left(bx + \frac{2n\pi}{4} \right).$$

$$(D^2 + a \cdot D + b) \cdot \sin \cdot pt = -p^2 \cdot \sin \cdot pt + ap \cdot \cos \cdot pt + b \cdot \sin \cdot pt,$$

$$\text{and } \frac{1}{D^2 + aD + b} \cdot \sin \cdot pt = \frac{\sin(pt - \alpha)}{\sqrt{(b - p^2)^2 + a^2 p^2}}, \quad \alpha = \tan^{-1} \frac{ap}{b - p^2}.$$

$$\text{If } u = b - p^2, b[21] \equiv z[A] : p[11] \rightarrow u \text{ or } (b - p^2)[6].$$

$$\text{If } v = ap, a[4] : p[9] \rightarrow v \text{ or } ap[8].$$

$$u[24] \nearrow u[4], v[8] \nearrow v[27]$$

$$u[3] : v[27] \rightarrow (u^2 + v^2)[24] \text{ and } \sqrt{u^2 + v^2}[25]$$

$$u[4] : v[8] \rightarrow v/u \text{ or } \alpha[9].$$

$$r \cdot e^{i\theta} \cdot A \cdot e^{ixt} = r \cdot A \cdot e^{i(xt + \theta)}, (r \cdot e^{i\theta})^n \cdot A \cdot e^{ixt} = r^n \cdot A \cdot e^{i(xt + n\theta)}.$$

These quantities are easily evaluated on nomogram 3.

Thus, nomogram 3 enables numerical solutions of a large number of differential equations to be evaluated.

10. Integration.

Many integrals may be evaluated using nomogram 3, for example:—

$$1. \quad u_n = \int x^n \cdot \cos \cdot bx \cdot dx = 1/b \cdot \sin \cdot bx \cdot x^n - n/b \cdot v_{n-1}.$$

$$v_n = \int x^n \cdot \sin \cdot bx \cdot dx = -1/b \cdot \cos \cdot bx \cdot x^n + n/b \cdot u_{n-1}.$$

$$\begin{aligned}
 2. \quad u_{m,n} &= \int \sin .^m \theta . \cos .^n \theta . d\theta, \\
 &= \frac{1}{m+n} \sin .^{m+1} \theta . \cos .^{n-1} \theta + \frac{n-1}{m+n} u_{m,n-2}, \\
 &= -\frac{1}{m+n} \sin .^{m-1} \theta . \cos .^{n+1} \theta + \frac{m-1}{m+n} u_{m-2,n}.
 \end{aligned}$$

$$\begin{aligned}
 3. \quad P &= \int e^{ax} . \cos . bx . dx = \frac{b \sin bx + a \cos bx}{a^2 + b^2} . e^{ax}. \\
 Q &= \int e^{ax} . \sin . bx . dx = \frac{a \sin bx - b \cos bx}{a^2 + b^2} . e^{ax}.
 \end{aligned}$$

11. Epilogue.

In this paper, an attempt has been made to show that nomography may be applied to any type of mathematical calculation, to the solution of differential equations, and to the study of errors of observation and laws. Much more could have been indicated to show that nomography is not a rough method of checking calculations, but that it is possible to work to four or five figure accuracy on a sheet of paper measuring about 50" by 30". The cost of printing the present paper is already high and the extra cost involved made the inclusion of a section on "Accuracy and Range" prohibitive.

I hope at some future time to indicate how nomography can very profitably be used quite early in mathematical and scientific studies. Why has this subject not received wider and more enthusiastic recognition? The great bar to the introduction of nomography into elementary work is of exactly the same nature as the bar to the introduction of the use of logarithms before tables of logarithms were available, or the use of squared paper when ready printed squared paper was unobtainable. If students had to calculate their own logarithms or draw their own squared paper, no one would use these most valuable aids to mathematical and scientific study. Moreover, a student can use logarithms or squared paper with immense profit long before he studies exponentials or analytical geometry. Similarly, nomograms will never come into general use until ready printed scales are available. This need not involve many different types of nomogram. Nomogram 3 of the present paper, suitably reconstructed and elaborated, will deal with most types of calculation. I have no doubt that once one suitable type of nomographic paper comes into use, many others will rapidly appear.

Those who are not in the habit of using graphical methods will probably ask why the publication of logarithmic paper and semi-logarithmic paper has not resulted in its more widespread use. There is more than one reason. The most important one is that it is possible to obtain a School Leaving Certificate and obtain a pass-with-credit in Elementary Mathematics without knowing anything about logarithmic plotting and, indeed, with no graphical work at all. "Graphs" form part of the syllabus for Elementary Mathematics, but there is no mention of graphical work for

a Higher Certificate. It is the engineer and the scientist who have been the pioneers of graphical methods. The training of engineers and scientists would lose little and gain much if the methods of pure and analytical geometry were heavily pruned and more prominence given to graphics (including projective geometry). For graphical methods to come to ultimate fruition, they must be taught early at every stage in the study of science and mathematics. Once a student gets into a graphical frame of mind, he will not be slow to use logarithmic paper and indeed, all manner of different rulings of paper will come into general use, *e. g.*, the harmonically ruled paper of fig. 3.4. Not only this, but transparent papers and papers with printed curves or other aids will also find ready and useful application.

One other matter should be mentioned. Some nomographic work can be carried out using ordinary squared paper. Conversion of units, testing experimental observations, $1/z = 1/x + 1/y$, etc., can be carried out using uniform scales. C. Dudley Langford, in a paper "A graphical method of solving problems on 'Rate of Work' and similar problems" (*Math. Gazette*, xxv. 267, Dec. 1941, p. 304), refers to the well-known type of problem in elementary arithmetic: "If one man takes 4 hours to do a piece of work and another 6 hours, how long will they take working together." He remarks, "Actually, children of 9 and even 8 love the simple first case" (he means that a third man does not join in later). A student who has tabulated $1/u$ and $1/v$ for a lens and then deduced f appreciates to the full the use of a reciprocal nomogram, and can use it after two minutes of explanation even if he has never used or heard of a nomogram before. Simple addition and subtraction nomograms could supplement the abacus in the earliest stages of arithmetic. The engineer and scientist, who from childhood has handled nomograms in the working of all manner of different types of problem, will be grateful for a method of study which eliminates needless laborious calculation and throws light on the numerical relationships between magnitudes which form the background of his work and the basis of his practical training.

II. Modern Philosophy of Science and some Problems of Classical Physics.

By G. BURNISTON BROWN, M.Sc., Ph.D.*

[Received October 13, 1942.]

IT has recently been observed in 'Nature' that the 'Philosophical Magazine' is once more becoming philosophical. A heading "Philosophy of Scientific Method" has appeared in 'Science Abstracts A.' In the August number of this journal Prof. W. Wilson, remarking that Prof.

* Communicated by the Author.

Dingle and I have strong philosophical bents, asks what practical difference it will make to physics if my suggestion for a new treatment of the theory of dimensions, based on considerations of a philosophical nature, are adopted. These facts make it clear that philosophy is entering physics and that there are some who do not feel that its entry is always justified. This is a challenge which, if the Editors permit, I shall attempt to answer in this paper.

At the outset it may be said that since the dimensional formulæ which follow from the theory I put forward are the simplest yet proposed, this is itself a practical advantage. But it is the new viewpoint which is of greater importance, and as Prof. Wilson says that he does not understand it, I shall attempt to clarify it by a discussion of its origins. And since it is notoriously dangerous to interpret the philosophy of others, this account will deal only with my own version of modern scientific philosophy and its application to some problems of classical physics, and the suggestion will be made, as modestly as may be, that the results are not without practical value to physicists in general.

Most unsolved problems of human knowledge, apart from those requiring more experiment, are either

- (A) Linguistic, *i. e.* due to insufficient care with regard to the meaning of words,
- or (B) Cultural, *i. e.* due to uncritical assumption of certain basic ideas peculiar to the culture of the period.

An example of the former is the problem of Achilles and the Tortoise. It appears to be a physical problem and to involve time and space, but the paradox rests on the misuse of the word "never," which is wrongly applied to the sum of a series of decreasing time intervals. An example of the latter is the difficulty the Greeks experienced in representing the diagonal of a square by a whole number when the sides were whole numbers of any unit. This is a mystery only for those living in a culture in which numbers are assumed to be the essence of all things and to represent tangible and visibly limited units, thus being only positive and whole. This is not a problem in Western European culture which, however, has others peculiar to itself, *e. g.*, due to the assumption of absolute time and space*.

Bearing the above observations in mind, consider the problem known to physicists as "B and H." The confusion is such that an international committee had to resort to a majority vote to decide whether B and H were different, and the result was the introduction of the bastard oersted. The secretary even went so far as to say that there are at least six definitions of the permeability μ and that it is impossible either to show that they are the same or different.

* For the underlying unity of all forms of expression of a culture see Spengler, 'The Decline of the West,' vol. 1 (Allen and Unwin), and also "Why do Archimedes and Eddington both get 10^{79} for the total number of Particles in the Universe?" 'Philosophy,' xv. p. 59 (1940).

Let us see if this problem comes under (A) or (B) above, or perhaps both. If it is linguistic we must inquire: what is the meaning of the statement that the magnetic force in the middle of a solid block of iron is, say, 8 dynes per unit pole? The eight is a number and it represents what Eddington calls a pointer reading. But how can we imagine a pointer reading obtained in the middle of a solid? Clearly we must have a test-particle (magnetic point-pole) and some arrangement like a spring-balance in order to give the pointer reading. But unless the test-particle is free to take up a position of equilibrium, the reading cannot be obtained. This involves a cavity for the act of measurement. Thus the consideration of cavities is necessary in order to give meaning to the words "force in a medium," and when we examine what are the minimum requirements for measuring the force on a point-pole and a current-element, we are led to cavities similar to those employed by Kelvin and Maxwell.⁽¹⁾ Thus the introduction of certain cavities, formerly arbitrary and mysterious, receives an explanation. I shall venture to call this (Benefit of Philosophy no. 1).

A similar consideration of the cavity necessary to establish the relation between the induced electromotive force and the rate of change of the magnetic flux, by means of the conservation of energy, shows clearly that it is the normal Induction which is involved and not the force. This is not often shown clearly in the textbooks. (B. of P. no. 2.)

Further, these considerations led to the detection of an error in Mosotti's formula connecting the dielectric constant K with the polarizability α , and to what I believe is the first correct relation between them, viz.,

$$K = 1 + 4\pi n\alpha,$$

where n is the number of doublets per cubic centimetre. If this is valid, it constitutes (B. of P. no. 3).

Once again, insistence on the meaning of the words "force in a medium" and therefore on cavities, resolves the confusion existing with regard to whether a current circuit and a magnetic shell which are equivalent in air are also equivalent in a medium (B. of P. no. 4).

So far we have only considered difficulties which can be overcome by concentration on the meaning of the words used. The fundamental problem about B and H comes under the heading (B), and is due to the ideas about space which characterize Western European culture*. In physics these ideas lead to the notion of a "field of force" which contains an inherent difficulty. This is because we have to suppose a magnetic field to act with one force and one direction on a pole and another force and another direction on a current element (the magnitude of the force is made the same in air by arbitrary definition of unit current) which amounts to having two different fields at one point. We are not content to treat the problem as one of interaction between different objects but prefer to consider it as interaction between different objects and space.

* See Spengler, *ibid.*

This leads to the question whether the two fields in space are really different, and whether one should be measured in gauss and the other in oersteds. If the Induction were treated as a measure of the interaction between a magnetic field and currents already flowing, or induced, in a conductor, then it would be reasonable to have a unit with a different name to signify that the force is measured per unit current per unit length of a conductor instead of per unit magnetic pole. But in the classical treatment both B and H are defined as forces per unit-pole, so that to have two units with different names is superfluous (B. of P. no. 5).

Epistemological considerations having focussed attention on the act of measurement⁽²⁾, I proceeded to examine whether there was any advantage in having symbols to represent operations of measurement and what relations, if any, would exist between them. This led to *A New Treatment of the Theory of Dimensions* in which L and T no longer stand for the unit (as in Maxwell's treatise), but for the *conventional operations of exact science for introducing numbers as symbols for the length and time intervals experienced in consciousness*⁽³⁾. Once M no longer stood for the unit gram it was possible to adopt the gravitational unit of mass, and as this is defined in terms of length and time only, M could be dropped altogether. Dropping M also agreed with a critical analysis of the way in which pointer readings are obtained. All dimensions can therefore be expressed in terms of L and T only, and fractional indices do not appear. This leads to the simplest scheme of dimensions yet proposed (B. of P. no. 6)*.

But the simplicity of the dimensional expressions, although of practical value to physicists, is the least interesting result of the application of philosophical criticism. I shall attempt now to explain some more important results of the new view of the philosophy of exact science which I have put forward recently^{(3), (4), (5), (6)}.

If the "whole subject matter of exact science consists of pointer readings and similar indications" (Eddington) then it is clear that exact science deals with numbers put into essentially-non-numerical Nature by the physicist. It follows that we should concentrate attention on the method of putting in the numbers rather than on the nature of the phenomena concerned. For instance, if we use a torsion balance for an electrical and a gravitational experiment, the *numbers* so obtained cannot possibly tell us the difference in nature between charged and uncharged matter (*e. g.*, that one can produce a "shock"). The laws of force derived from such experiments are relations between the numbers obtained and so, again, cannot exhibit any difference in nature between electricity and matter. We can, of course, employ extra symbols such as $+$ and $-$ in order to indicate the difference between attraction and repulsion which differentiates electricity from gravitation, but these are

* For certain technical purposes it is, no doubt, necessary to retain a separate dimension for mass.

not numbers. As Eddington has pointed out, in exact science e the electric charge is a measure of the mutual interaction of two charges. It is not something you rub off glass with a silk duster: it is a number produced as the result of measurement. So also, of course, is M the mass, and m the magnetic pole. Considerations of this kind led me to suggest that the dimensional formulæ for mass, electric charge, and magnetic pole should be identical.

Another point of great interest arose in connexion with the question: if numbers are introduced according to more than one system, what relation must hold between the numbers expressing the measurement of one and the same physical quantity in these different systems? For example, if a student wishes to measure an Inductance on the E.M. system, it is sufficient to give him a scale: if another student wants to measure the same Inductance on the E.S. system we have to hand him out a clock as well as a scale. Now if the two students measure all the Inductances in the laboratory, we should expect the numbers which they produce in each case to bear a constant ratio to one another, otherwise they cannot be measuring the same quantity (the original reasons for believing that both measure the same quantity are theoretical). But a constant ratio between measurements made in one case with a scale and in the other with a scale and clock, suggest that there must be a length measurement equivalent to a time measurement, and vice versa, and that the equivalence will be expressed by a universal constant.

This led to a critical examination of length and time measurements, and to the conclusion that a length measurement always involves a minimum time interval and that the number so obtained ought to be regarded as the measure of a space-time interval. Dimensionally this was symbolized by

$$L=cT$$

where, on the c.g.s. system, c is very close to 3×10^{10} cm. per sec. A length measurement can, of course, occupy any time, but it cannot take less than l/c secs. where l is the length in cm. and c is a universal constant. This is somewhat analogous to the quantum principle, where we may have any energy associated with a given frequency, but not less than $h\nu$, where ν is the frequency and h is a universal constant.

With this "rate of exchange" between length and time measurements, we can proceed to examine the equivalence or otherwise of different ways of measuring the same quantity. The measurement of an Inductance on the E.S. system, for instance, is symbolized by $L^{-1}T^2$. Substituting from $L=cT$ this is equivalent to L/c^2 . But L represents the measurement of an Inductance on the E.M. system, so that if we take the ratio of the results of an E.S. measurement and an E.M. measurement we are bound to get c^2 . Since the measurement of all physical quantities can be expressed in terms of L and T , it follows that if the same quantity is measured according to two systems both of which are founded on one original measure-system (that of mechanics in the

case of the E.S. and E.M. systems), the measurements must be represented by L and T raised to different powers in the two cases. If these are equivalent, they must be interchangeable by means of the relation $L=cT$, so that c , or some power of it, must be the ratio of numbers expressing the value of a quantity measured on different systems.

Thus the appearance of c , the velocity of light, in the results of electrical experiments in which nothing optical is observed or measured is explained. From this point of view, we do not look upon such results as showing that light is electromagnetic in nature, as the classical physicists did, but as evidence supporting the theory which led to the belief that the different measuring operations were equivalent, *i. e.*, that they measured the same physical quantity. This theory is, of course, based on the hypothesis that a current is equivalent to electric charge in motion.

This new theory of dimensions, therefore, leads to the general result that if numbers are introduced according to the measure-convention of mechanics *in two different ways*, then the ratio of the measurements of the same physical quantity on the two systems must be a power of c the velocity of light (or "constant of interaction"). (B. of P. no. 7.) The theory also allows us to go further and examine the equivalence between measures of *different* physical quantities. This leads at once to a very simple derivation of the mass-energy relation (B. of P. no. 8). The theory does not predict whether mass can "turn into" energy, but asserts that if such an interchange is found in Nature, the numbers obtained by conventional measuring processes will bear a certain relationship, *viz.*, $energy=c^2 mass$. It does, however, predict that mass cannot "turn into," *e. g.*, action.

Finally, the fact that science deals with pointer readings and that these may be reduced to those of length and time only, explains how it is possible to reduce the results of metrical physics to nothing but the properties of a *space-time* continuum (B. of P. no. 9).

The emphasis developed in this "philosophy of exact science" on measuring operations rather than on the "nature" of the subject, leads to the following expression of the aims of exact science⁽⁶⁾. For the sake of consistency and to avoid unnecessary confusion, we should introduce numbers into all branches of science according to the same system. Since the system of measurement of mechanics was historically the first and has been used for introducing numbers into the study of electricity and magnetism, the aim of exact science should be to found all measurements on those of mechanics.

The only subject of physics now offering itself for the fruitful pursuit of this aim is heat. In a recent paper⁽⁶⁾, I have introduced numbers into the study of heat by commencing with mechanical measurement (pressure) instead of arbitrary thermometer readings. In this way a dynamical scale of temperature independent of the properties of any particular substance, arises naturally and clearly, and the erg becomes the natural unit of quantity of heat. Following from this, calorimetry can be founded directly on the principle of conservation of energy,

instead of on a (usually undisclosed) historical background of an indestructible caloric fluid. The connexion between heat measured in calorimetry and heat measured directly in mechanical terms can be obtained simply without recourse to Carnot's Cycle, the obscurities of which are largely due to metaphysical difficulties about the "nature" of heat. Here again, therefore, I submit, are problems and obscurities of classical physics which fade away when this new outlook on exact science is adopted.

The above account is extremely brief and is only meant as a guide to those who have read recent papers of mine and cannot see "what they are about," or who, like Prof. Wilson, are doubtful as to whether they have any practical value. The underlying ideas of this philosophy of exact science can be summed up as follows:—

Exact science is that part of science which is concerned with the introduction of numbers into essentially-non-numerical Nature. Concentration of attention on acts of measurement is therefore essential in order to give physical meaning to the numbers *put in*. It may also throw some light on the numbers that *come out*. For example, it explains why the ratio of E.M. to E.S. quantities is always a power of c and why energy must be c^2 times mass. In considerations of this kind it is therefore necessary to examine what the physicist actually does, and that is the reason for my proposal for a new treatment of the theory of dimensions so as to form a calculus of *operations*.

References.

- (1) "A New Treatment of Electric and Magnetic Induction." Proc. Phys. Soc. lii. p. 577 (1940).
- (2) "Fundamentals of Classical Electric and Magnetic Theory." 'Nature,' cxlv. p. 789 (1940).
- (3) "A New Treatment of the Theory of Dimensions." Proc. Phys. Soc. liii. p. 418 (1941).
- (4) "The Philosophy of Physical Science." 'Nature,' cxlviii. p. 504 (1941).
- (5) "A Note on Prof. Wilson's 'Dimensions of Physical Quantities.'" Phil. Mag. xxxi. p. 367 (1942).
- (6) "A Dynamical Treatment of the Elements of Heat." Phil. Mag. xxxi. p. 543 (1942).

III. On Some Infinite Sets of Numbers.

By LUDWIK SILBERSTEIN, PH.D.*

[Received August, 18, 1942.]

CONSIDER the set of numbers u_1, u_2, u_3, \dots defined by the requirement that, starting with u_3 , each number shall be the *arithmetic mean* of the preceding two,

$$u_n = \frac{1}{2}(u_{n-1} + u_{n-2}). \quad \dots \quad (1)$$

* Communicated by the Author.

Try to satisfy this requirement by assuming u_n proportional to α^n , where α is a constant. Then (1) gives

$$x^2 = \frac{1}{2}(\alpha + 1),$$

a quadratic for α . Its two roots are $\alpha_1 = 1$, $\alpha_2 = -\frac{1}{2}$. Thus the general solution of (1) is

$$u_n = c_1 \pm \frac{c_2}{2^n}, \quad \dots \dots \dots (1.1)$$

where c_1 , c_2 are arbitrary constants and the signs \pm hold for n even
odd.

The two constants will be determined by any pair of constituents of the set. Let u_1 , u_2 be this pair. Then $c_1 - \frac{1}{2}c_2 = u_1$, $c_1 + \frac{c_2}{4} = u_2$, whence $c_1 = \frac{2}{3}u_2 + \frac{1}{3}u_1$, $c_2 = (u_2 - u_1)$, and (1.1) becomes

$$u_n = \frac{1}{3}(u_1 + 2u_2) \pm \frac{u_2 - u_1}{3 \cdot 2^{n-2}} \quad \dots \dots \dots (1.2)$$

In the trivial case $u_2 = u_1$ the variable term vanishes and $u_n = u_1 = \text{constant}$, as could be expected. For any $u_2 \neq u_1$ the set is oscillatory, with ever decreasing amplitude; it forms what is technically called a "sequence." The limit of (or the number defined by) this sequence is

$$u_\infty = \frac{u_1 + 2u_2}{3} \quad \dots \dots \dots (1.3)$$

It seems remarkable that, whatever the values of the first two constituents, the limit is their weighted mean, u_1 being given the weight *one* and u_2 the weight *two*.

Consider now the somewhat more complicated requirement that each number of the set (beginning with the fourth) shall be the arithmetic mean of the preceding three,

$$u_n = \frac{1}{3}(u_{n-1} + u_{n-2} + u_{n-3}). \quad \dots \dots \dots (2)$$

To satisfy this requirement assume again u_n proportional to α^n . Then

$$\alpha^3 = \frac{1}{3}(\alpha^2 + \alpha + 1),$$

a cubic for α . One root is $\alpha = 1$, leaving for the remaining two the quadratic

$$\alpha^2 + \frac{2}{3}\alpha + \frac{1}{3} = 0.$$

The complex conjugate pair of its roots is $\alpha = -\frac{1}{3} \pm \frac{\sqrt{2}}{3}i$,

$$\left. \begin{aligned} \text{or} \quad \alpha_1 &= \frac{1}{\sqrt{3}} e^{i\theta}, & \alpha_2 &= \frac{1}{\sqrt{3}} e^{-i\theta}, \\ \text{where} \quad \sin \theta &= \frac{\sqrt{2}}{3}, & \cos \theta &= -\frac{1}{\sqrt{3}}. \end{aligned} \right\} \quad \dots \dots \dots (2.1)$$

The general solution of (2) is

$$u_n = c + c_1 \alpha_1^n + c_2 \alpha_2^n, \quad \dots \quad (2.2)$$

where c, c_1, c_2 are arbitrary constants, which can be determined from any given triad of constituents. Let us express them again in terms of the first triad, u_1, u_2, u_3 , i. e., from the three linear equations—

$$\begin{aligned} c + \alpha_1 c_1 + \alpha_2 c_2 &= u_1, \\ c + \alpha_1^2 c_1 + \alpha_2^2 c_2 &= u_2, \\ c + \alpha_1^3 c_1 + \alpha_2^3 c_2 &= u_3. \end{aligned}$$

Since, by (2.1), $\alpha_1 \alpha_2 = \frac{1}{3}$, $\alpha_1 + \alpha_2 = -\frac{2}{3}$, $\alpha_2 - \alpha_1 = -\frac{2\sqrt{2}i}{3}$, the determinant of this system is readily found to be $-\frac{4\sqrt{2}}{9}i$ and

$$\begin{aligned} c &= \frac{1}{6}(u_1 + 2u_2 + 3u_3), \\ c_1 &= \frac{9i}{4\sqrt{2}} \alpha_2(1 - \alpha_2)\{\alpha_2 u_1 - (1 + \alpha_2)u_2 + u_3\}, \\ c_2 &= \frac{-9i}{4\sqrt{2}} \alpha_1(1 - \alpha_1)\{\alpha_1 u_1 - (1 + \alpha_1)u_2 + u_3\}. \end{aligned}$$

Substituting these values of the coefficients in (2.2) one finds, after some straightforward reductions,

$$\begin{aligned} u_n &= \frac{1}{6}(u_1 + 2u_2 + 3u_3) - \frac{3}{4}\sqrt{\frac{2}{3}}\left\{(u_1 - u_2)\left[\sin(n-2)\theta - \frac{1}{\sqrt{3}}\sin(n-3)\theta\right] \right. \\ &\quad \left. + (u_3 - u_2)[\sqrt{3}\sin(n-1)\theta - \sin(n-2)\theta]\right\}. \quad \dots \quad (2.3) \end{aligned}$$

This is the required solution. It is readily verified to give u_1, u_2, u_3 for $n=1, 2, 3$.

If $u_1 = u_2 = u_3$, the variable terms disappear and we are left with $u_n = u_1 = \text{constant}$, as could be expected. In all other cases (2.3) represents a sequence oscillating in ever-decreasing amplitudes about its limit

$$u_\infty = \frac{1}{6}(u_1 + 2u_2 + 3u_3), \quad \dots \quad (2.4)$$

which again is the “weighted” mean of the first three constituents, u_1 being given the weight *one*, u_2 the weight *two*, and u_3 the weight *three*.

This remarkable property is a natural extension of that noted in the first case. It holds also for a set in which every constituent is the simple arithmetical mean of its four, or its five, etc., predecessors. (A formal proof may be left to the reader.) It is chiefly in view of this property of the first dyad or triad, etc., in moulding the limit that this subject has seemed to deserve some attention.

129 Seneca Parkway,
Rochester, N.Y.
July 6, 1942.

IV. *Notes on the Reversion of a Series.*

By W. G. BICKLEY, D.Sc., Imperial College of Science and Technology,
and J. C. MILLER, Ph.D., University of Liverpool *.

[Received August 4, 1942.]

I.

By W. G. BICKLEY.

IN obtaining the McMahon expansions for the zeros of Bessel functions, series must be reversed of a type which is not covered by the general formulæ given by van Orstrand (Phil. Mag. (6) xix. pp. 366-376, 1910). The principle of one of the methods given is, however, applicable to these series.

If the relation

$$y = x + a/x + b/x^3 + c/x^5 + \dots \quad (1)$$

becomes, upon reversion,

$$x = y + A/y + B/y^3 + C/y^5 + \dots \quad (2)$$

then, on differentiating (2) with respect to x ,

$$1 = y'(1 - A/y^2 - 3B/y^4 - \dots - (2n+1)Y_n/y^{2n+2} \dots).$$

Multiply by y^{2n+1} and obtain

$$y^{2n+1} = y'(y^{2n+1} - Ay^{2n-1} - \dots - (2n+1)Y_n/y - \dots). \quad (3)$$

Now if the right-hand side of (3) be expressed in terms of x the only term which can contain a multiple of $1/x$ is that involving y'/y . It follows that

$-(2n+1)Y_n$ is the coefficient of $1/x$ in the expansion of y^{2n+1} in powers of x .

In this manner it is found that

$$A = -a$$

$$B = -(b + a^2)$$

$$C = -(c + 4ab + 2a^3)$$

$$D = -(d + 6ac + 3b^2 + 15a^2b + 5a^4)$$

$$E = -(e + 8ad + 8bc + 28ab^2 + 28a^2c + 56a^3b + 14a^5)$$

$$F = -(f + 10ae + 10bd + 5c^2 + 45a^2d + 90abc + 15b^3 + 120a^3c + 180a^2b^2 + 210a^4b + 42a^6)$$

* Communicated by the Authors.

$$G = -(g + 12af + 12be + 12cd + 66a^2e + 132abd + 66ac^2 \\ + 66b^2c + 220a^3d + 220ab^3 + 660a^2bc + 495a^4c \\ + 990a^3b^2 + 792a^5b + 132a^7)$$

The reciprocity of (1) and (2) makes a complete and independent check of these formulæ readily possible. To detect missing terms, the sums of the coefficients are useful. To obtain them, put in (1) $a=b=c=\dots=1$; then

$$y = x^3/(x^2 - 1),$$

so that

$$\begin{aligned} -(2n+1)Y_n &= \text{coeff. of } 1/x \text{ in } \{x/(1-1/x^2)\}^{2n+1} \\ &= \text{coeff. of } x^{-(2n+2)} \text{ in } (1-1/x^2)^{-(2n+1)} \\ &= \frac{(2n+1)(2n+3)\dots(3n+1)}{(n+1)!}. \end{aligned}$$

It follows that the sum of the coefficients in Y_n is

$$\frac{(3n+1)!}{(n+1)!(2n+1)!}.$$

These tests have, of course, been applied.

II.

By J. C. P. MILLER.

The method indicated in the preceding note makes it very simple to find terms in the reversed series; to these terms the simplification described by Bleick (Phil. Mag. (7) xxxiii. pp. 637-8, 1942) is applicable.

The notation of the preceding note is used, with the addition of

$$\begin{aligned} \alpha &= Aa & \gamma &= Ca + Bb + Ac \\ \beta &= Ba + Ab & \delta &= Da + Cb + Bc + Ad \end{aligned}$$

etc. The results obtained are as follows :—

$$\begin{aligned} A+a &= 0 & D+d &= 3\gamma + \alpha^2 \\ B+b &= \alpha & E+e &= 4\delta + 6\alpha\beta \\ C+c &= 2\beta & F+f &= 5\epsilon + 15\alpha\gamma + 5\beta^2 + 7\alpha^3 \\ G+g &= 6\zeta + 31\alpha\delta + 19\beta\gamma + 73\alpha^2\beta \\ H+h &= 7\eta + 56\alpha\epsilon + 35\beta\delta + 14\gamma^2 + 238\alpha^2\gamma + 154\alpha\beta^2 + 110\alpha^4 \end{aligned}$$

The results have been checked by putting $a=b=c=\dots=1$, and obtaining the values of A, B, C, \dots in succession. These were found to be $-(3n+1)!/\{(n+1)!(2n+1)!\}$ as before.

V. Formulæ Relating to Bessel Functions of Moderate or Large Argument and Order.

By W. G. BICKLEY, D.Sc.,
Imperial College of Science and Technology*.

[Received August 4, 1942.]

1. SEARCH of extant literature for formulæ to aid in the computation of the zeros of the derivatives of Bessel functions revealed large gaps in the available information, and it became necessary to derive the desired formulæ, particularly those corresponding to the various asymptotic expansions for the functions themselves. The investigation brought to light further interconnexions between formulæ for the derivatives and those for the functions, and some new results for the functions; it has therefore seemed desirable to give a comprehensive set of formulæ, even if this involves the repetition of a few well-known ones.

The aim is purely utilitarian, in the sense that the formulæ are needed as bases for computation, and that explicit developments to a number of terms, rather than analytical proofs of the existence of such developments, are therefore essential. Elementary methods have been used, and such processes as the term-by-term differentiation of asymptotic expansions freely employed.

The notation used is, generally, that of Watson's 'Theory of Bessel Functions,' although additions and occasional modifications have been made. In particular, n has not been reserved solely for integral orders. References to pages of Watson's treatise are given thus (W 228).

We are primarily concerned with the functions $J_n(x)$ and $Y_n(x)$, but conciseness will often be achieved by the use of one or other of the combinations

$$H_n^{(1)}(x) = J_n(x) + iY_n(x)$$

or

$$\mathcal{E}_n(x) = J_n(x) \cos \alpha - Y_n(x) \sin \alpha,$$

where α is independent both of n and of x . Where no confusion can arise, order, argument, or both may be omitted, so that J alone may represent $J_n(x)$.

2. It is known that (W 199)

$$\begin{aligned} J_n(x) &= \sqrt{(2/\pi x)} \{ P_n(x) \cos \theta_n - Q_n(x) \sin \theta_n \} \\ Y_n(x) &= \sqrt{(2/\pi x)} \{ Q_n(x) \cos \theta_n + P_n(x) \sin \theta_n \} \end{aligned} \quad \dots \quad (1)$$

* Communicated by the Author.

where, if $x \gg n$, and also $\gg 1$, and $\mu = 4n^2$,

$$\left. \begin{aligned} P_n(x) &\sim 1 - \frac{(\mu-1)(\mu-9)}{2! (8x)^2} + \frac{(\mu-1)(\mu-9)(\mu-25)(\mu-49)}{4! (8x)^4} + \dots \\ Q_n(x) &\sim \frac{(\mu-1)}{8x} - \frac{(\mu-1)(\mu-9)(\mu-25)}{3! (8x)^3} + \dots \end{aligned} \right\} \quad (2)$$

and

$$\theta_n = x - \frac{1}{2}(n + \frac{1}{2})\pi.$$

It follows that

$$H_n^1(x) = \sqrt{(2/\pi x)} \{P_n(x) + iQ_n(x)\} e^{i\theta_n}. \quad (3)$$

$$\mathcal{E}_n(x) = \sqrt{(2/\pi x)} \{P_n(x) \cos(\theta_n + \alpha) - Q_n(x) \sin(\theta_n + \alpha)\}. \quad (4)$$

Similarly we write

$$\left. \begin{aligned} J_n'(x) &= \sqrt{(2/\pi x)} \{-R_n(x) \sin \theta_n - S_n(x) \cos \theta_n\} \\ Y_n'(x) &= \sqrt{(2/\pi x)} \{-S_n(x) \sin \theta_n + R_n(x) \cos \theta_n\} \end{aligned} \right\} \quad (5)$$

where

$$\left. \begin{aligned} R_n(x) &= P_n(x) + Q_n'(x) - \frac{1}{2}x^{-1}Q_n(x) \\ &\sim 1 - \frac{(\mu-1)(\mu+15)}{2! (8x)^2} + \frac{(\mu-1)(\mu-9)(\mu-25)(\mu+63)}{4! (8x)^4} - \dots \\ S_n(x) &= Q_n(x) - P_n'(x) + \frac{1}{2}x^{-1}P_n(x) \\ &\sim \frac{(\mu+3)}{8x} - \frac{(\mu-1)(\mu-9)(\mu+35)}{3! (8x)^3} + \dots \end{aligned} \right\} \quad (6)$$

The general term of the $R-S$ series is

$$\frac{(\mu-1) \dots (\mu-2r-3^2)(\mu+4r^2-1)}{r! (8x)^r}.$$

We have, also,

$$H_n^{1'}(x) = \sqrt{(2/\pi x)} \{R_n(x) + iS_n(x)\} e^{i(\theta_n + \frac{1}{2}\pi)}. \quad (7)$$

$$\mathcal{E}_n(x) = \sqrt{(2/\pi x)} \{-R_n(x) \sin(\theta_n + \alpha) - S_n(x) \cos(\theta_n + \alpha)\}. \quad (8)$$

3. Let

$$\left. \begin{aligned} U_n(x) &= J_n^2(x) + Y_n^2(x) = M_n^2(x) \\ &= (2/\pi x) \{P_n^2(x) + Q_n^2(x)\} \end{aligned} \right\} \quad (9)$$

and

$$\left. \begin{aligned} V_n(x) &= J_n'^2(x) + Y_n'^2(x) = N_n^2(x) \\ &= (2/\pi x) \{R_n^2(x) + S_n^2(x)\} \end{aligned} \right\} \quad (10)$$

Then

$$\begin{aligned} dU/dx &= 2(JJ' + YY') \\ d^2U/dx^2 &= 2(J'^2 + Y'^2 + JJ'' + YY'') \\ &= 2V - 2 \left\{ \frac{JJ' + YY'}{x} + \left(1 - \frac{n^2}{x^2}\right)(J^2 + Y^2) \right\} \\ &= 2V - U'/x - 2(1 - n^2/x^2)U, \end{aligned}$$

or

$$2x^2V = x^2U'' + xU' + 2(x^2 - n^2)U. \quad (11)$$

Also

$$\begin{aligned} dV/dx &= 2(J'J'' + Y'Y'') \\ &= -2 \left\{ \frac{J'^2 + Y'^2}{x} + \left(1 - \frac{n^2}{x^2}\right)(JJ' + YY') \right\} \end{aligned}$$

or
$$\frac{d}{dx}(x^2 V) = -(x^2 - n^2) \frac{dU}{dx}. \quad (12)$$

Eliminating V between (11) and (12) yields

$$x^2 U''' + 3x U'' + U' + 4\{(x^2 - n^2)U' + xU\} = 0. \quad (13)$$

Since

$$U \sim 2/\pi x,$$

write

$$U_n(x) = \frac{2}{\pi} \sum \frac{a_{2r-1}}{x^{2r-1}},$$

where $a_1 = 1$. By substitution in the differential equation (13) and equating to zero the coefficients of powers of x , the recurrence formula

$$\{\mu - (2r-1)^2\}(2r-1)a_{2r-1} = 4 \cdot 2r a_{2r+1} \quad (14)$$

is obtained; hence

$$U_n(x) \sim \frac{2}{\pi x} \left\{ 1 + \frac{1}{2} \frac{(\mu-1)}{(2x)^2} + \frac{1 \cdot 3}{2 \cdot 4} \frac{(\mu-1)(\mu-9)}{(2x)^4} + \dots \right\}. \quad (15)$$

Equation (11) now gives

$$V_n(x) \sim \frac{2}{\pi x} \left\{ 1 - \frac{1}{2} \frac{(\mu-3)}{(2x)^2} - \frac{1 \cdot 1}{2 \cdot 4} \frac{(\mu-1)(\mu-45)}{(2x)^4} \dots \right\}, \quad (16)$$

the general terms of the series being

$$-\frac{1 \cdot 1 \cdot 3 \dots (2r-3)}{2 \cdot 4 \cdot 6 \dots 2r} \frac{(\mu-1)(\mu-9) \dots (\mu-2r-3)^2 \{\mu - (2r+1)(2r-1)^2\}}{(2x)^{2r}}. \quad (16 a)$$

It may readily be verified that the expansions (15) and (16) are consistent with the relations between U P Q and V R S contained in (9) and (10) and the expansions (2) and (6).

4. The Wronskian identity (W 76)

$$JY' - J'Y = 2/\pi x \quad (17)$$

gives several expressions, which are useful in checking algebra or computations.

First, if we insert in (17) the expressions for J , J' , Y and Y' given in (1) and (5), we obtain

$$PR + QS = 1. \quad (18)$$

Again, the identity

$$(J^2 + Y^2)(J'^2 + Y'^2) = (JJ' + YY')^2 + (JY' - J'Y)^2$$

yields the relation

$$UV = \frac{1}{4}U'^2 + 4/\pi^2 x^2. \quad (19)$$

Other consequences of (17) will be encountered later.

5. A form alternative to (1) has been used by Meissel (W 229), namely

$$\left. \begin{aligned} J_n(x) &= M_n(x) \cos \psi_n \\ Y_n(x) &= M_n(x) \sin \psi_n \end{aligned} \right\}, \quad (20)$$

so that

$$H_n^{(1)}(x) = M_n(x) e^{i\psi_n}, \quad (21)$$

$$\mathcal{E}_n(x) = M_n(x) \cos(\psi_n + \alpha). \quad (22)$$

Comparison with (1), (3), (4) and (9)—in the last of which this use of M_n was anticipated—yields

$$M^2 = U = (2/\pi x)(P^2 + Q^2), \quad (9 \text{ bis})$$

$$\psi = \theta + \tan^{-1}(Q/P). \quad (23)$$

Also, direct from (20),

$$\psi = \tan^{-1}(Y/J),$$

so that

$$\frac{d\psi}{dx} = \frac{JY' - J'Y}{J^2 + Y^2} = \frac{2/\pi x}{M^2} = \frac{2}{\pi x U}. \quad (24)$$

Since $\psi \rightarrow \theta$ as $x \rightarrow \infty$, expansion of $2/\pi x U$ and term-by-term integration leads to the result (W 506) *

$$\psi_n \sim x - \frac{1}{2}(n + \frac{1}{2})\pi + \frac{(\mu - 1)}{8x} + \frac{(\mu - 1)(\mu - 25)}{384x^3} + \dots \quad (25)$$

A zero of \mathcal{E} is given by

$$\psi + \alpha = \frac{1}{2}(2s - 1)\pi. \quad (26)$$

so that

$$\beta = \frac{1}{4}(2n + 4s - 1)\pi - \alpha \sim x + \frac{\mu - 1}{8x} + \dots \quad (27)$$

Reversion of this series gives McMahon's well-known expansion for the zeros of the Bessel functions (W 506)

$$x \sim \beta \left\{ 1 - \frac{\mu - 1}{8\beta^2} - \frac{(\mu - 1)(7\mu - 31)}{6(8\beta^2)^2} - \frac{(\mu - 1)(83\mu^2 - 982\mu + 3779)}{30(8\beta^2)^3} \right. \\ \left. - \frac{(\mu - 1)(6949\mu^3 - 53855\mu^2 + 1585743\mu - 6277327)}{840(8\beta^2)^4} \right. \\ \left. - \frac{(\mu - 1)(70197\mu^4 - 2479316\mu^3 + 48010494\mu^2 - 512062548\mu + 2092163573)}{2520(8\beta^2)^5} - \dots \right\}. \quad (28)$$

6. In like manner (partly anticipated in (10)) we write

$$\left. \begin{aligned} J_n'(x) &= -N_n(x) \sin \chi_n \\ Y_n'(x) &= N_n(x) \cos \chi_n \end{aligned} \right\}, \quad (29)$$

so that

$$H_n^{(1)'}(x) = N_n(x) e^{i(\chi_n + \frac{1}{2}\pi)}, \quad (30)$$

$$\mathcal{E}_n(x) = -N_n(x) \sin(\chi_n + \alpha). \quad (31)$$

Comparison with (5), (7), (8) and (10) shows that

$$N^2 = V = (2/\pi x)(R^2 + S^2), \quad (10 \text{ bis})$$

$$\chi = \theta + \tan^{-1}(S/R),$$

so that

$$\chi \rightarrow \theta \text{ as } x \rightarrow \infty.$$

Also, direct from (29)

$$\chi = -\tan^{-1}(J'/Y').$$

* Note that ψ as used here is not the same as the ψ of (W 506).

so that
$$\frac{d\chi}{dx} = \frac{J'Y'' - Y'J''}{J'^2 + Y'^2} = \left(1 - \frac{n^2}{x^2}\right) \frac{2}{\pi x V} \quad \dots \quad (33)$$

upon use of Bessel's differential equation.

Utilizing (16) to develop $(2/\pi x V)$ in descending powers of x , and then integrating (33) and determining the constant of integration from the asymptotic value of χ , we find

$$\begin{aligned} \chi_n = & x - \frac{1}{2}(n + \frac{1}{2})\pi + \frac{\mu + 3}{8x} + \frac{\mu^2 + 46\mu - 63}{384x^3} + \frac{\mu^3 + 185\mu^2 - 2053\mu + 1899}{5120x^5} \\ & + \frac{5\mu^4 + 2380\mu^3 - 88354\mu^2 + 629196\mu - 543483}{229376x^7} \\ & + \frac{7\mu^5 + 6825\mu^4 - 596554\mu^3 + 13351010\mu^2 - 85011885\mu + 72251109}{2359296x^9} \\ & + \frac{21\mu^6 + 36498\mu^5 - 6201349\mu^4 + 313105980\mu^3 - 5975156869\mu^2}{46137344x^{11}} \\ & + \dots \quad (34) \end{aligned}$$

If $\mathcal{E}'_n = 0,$

$$\chi_n + \alpha = (s - 1)\pi, \quad \dots \quad (35)$$

so that $\beta = \frac{1}{4}(2n + 4s - 3)\pi - \alpha \sim x + \frac{\mu + 3}{8x} + \frac{\mu^2 + 46\mu - 63}{384x^3} + \dots \quad (36)$

Reversion of this series (general formulæ for this type of series are given in the preceding note) yields the McMahon expansion for the zeros of the derivatives of the Bessel functions (W 507)

$$\begin{aligned} x = \beta \left\{ 1 - \frac{\mu + 3}{8\beta^2} - \frac{7\mu^2 + 82\mu - 9}{6(8\beta^2)^2} - \frac{83\mu^3 + 2075\mu^2 - 3039\mu + 3537}{30(8\beta^2)^3} \right. \\ - \frac{6949\mu^4 + 296492\mu^3 - 1248002\mu^2 + 7414380\mu - 5853627}{840(8\beta^2)^4} \\ - \frac{70197\mu^5 + 4535387\mu^4 - 38051230\mu^3 + 527973862\mu^2 - 2491515495\mu + 2014126479}{2520(8\beta^2)^5} \\ \left. - \dots \right\} \quad (37) \end{aligned}$$

7. We can recover, in terms of M , N , ψ and χ , some of our previous results. Since

$$H_n^{-1}(x) = Me^{i\psi}, \quad \dots \quad (21 \text{ bis})$$

then $H_n^{-1'}(x) = (iM\psi' + M')e^{i\psi}, \quad \dots \quad (38)$

and $H_n^{-1''}(x) = (iM\psi'' + 2iM'\psi' + M'' - M\psi'^2)e^{i\psi}. \quad \dots \quad (39)$

Bessel's differential equation yields

$$x^2(M'' - M\psi'^2) + xM' + (x^2 - n^2)M + i\{x^2(M\psi'' + 2M'\psi') + xM\psi'\} = 0, \quad \dots \quad (40)$$

and the real and imaginary parts must separately vanish. From the imaginary part we find

$$\frac{d}{dx}(xM^2\psi')=0,$$

so that

$$\psi'=A/xM^2.$$

The constant A is determined from the knowledge that as $x \rightarrow \infty$, $\psi' \rightarrow 1$ and $xM^2 \rightarrow 2/\pi$. Hence

$$\psi' = \frac{2/\pi x}{M^2} = \frac{2}{\pi x U}. \quad \dots \dots \dots (24 \text{ bis})$$

Using this result, the real part of (40) gives

$$x^2 M'' + xM' + (x^2 - n^2)M = 4/\pi^2 M^3. \quad \dots \dots \dots (41)$$

Since

$$2MM' = U', \quad 2(MM'' + M'^2) = U'', \quad \dots \dots \dots (42)$$

and in virtue of these, (41) becomes (upon doubling)

$$x^2(UU'' - \frac{1}{2}U'^2) + xUU' + 2(x^2 - n^2)U^2 = 8/\pi^2.$$

Differentiating with respect to x and dividing by U we recover

$$x^2 U''' + 3xU'' + U' + 4\{(x^2 - n^2)U' + xU\} = 0. \quad \dots \dots (13 \text{ bis})$$

Again, from (30) and (38),

$$H_n^{(1)}(x) = iNe^{ix} = (iM\psi' + M')e^{i\psi}, \quad \dots \dots \dots (43)$$

and from this we derive :

(i), by equating moduli

$$N^2 = M^2\psi'^2 + M'^2,$$

or, by (24) and slight rearrangement,

$$M^2 N^2 = 4/\pi^2 x^2 + (MM')^2, \quad \dots \dots \dots (44)$$

which is equivalent to (19) ;

(ii), by equating phases, again using (24),

$$\chi = \psi - \tan^{-1}(\frac{1}{2}\pi x M M'), \quad \dots \dots \dots (45)$$

so that

$$\chi' = \psi' - \frac{\frac{1}{2}\pi(MM' + xMM'' + xM'^2)}{1 + \frac{1}{4}\pi^2 x^2 M^2 M'^2}.$$

By the use of (11), (42) and (44) this reduces to

$$\frac{d\chi}{dx} = \left(1 - \frac{n^2}{x^2}\right) \frac{2}{\pi x V}. \quad \dots \dots \dots (33 \text{ bis})$$

We also note that, in terms of M , N , ψ and χ , the Wronskian relation (17) becomes

$$MN \cos(\psi - \chi) = 2/\pi x, \quad \dots \dots \dots (46)$$

a relation also deducible from (44) and (45).

8. A comparatively simple expression exists for the value of the derivative of a cylinder function at any of its zeros. Since

$$\mathcal{E}_n(x) = M \cos(\psi + \alpha),$$

then, at a zero,

$$\psi + \alpha = (s - \frac{1}{2})\pi.$$

But

$$\mathcal{E}_n'(x) = M' \cos(\psi + \alpha) - M\psi' \sin(\psi + \alpha),$$

so that at a zero

$$\mathcal{E}_n'(x) = (-)^s M\psi' = (-)^s 2/\pi x M = (-)^s 2/\pi x \sqrt{U}. \quad (47)$$

An equally simple expression gives the value of the function at the zeros of the derivative, i. e., a maximum or minimum of $\mathcal{E}_n(x)$. Since

$$\mathcal{E}_n'(x) = -N \sin(\chi + \alpha),$$

a zero of $\mathcal{E}_n'(x)$ is given by

$$\chi + \alpha = (s - 1)\pi.$$

At this zero

$$\begin{aligned} \mathcal{E}_n(x) &= M \cos(\psi + \alpha) \\ &= M \cos(\psi - \chi + \chi + \alpha) \\ &= (-)^{s-1} M \cos(\psi - \chi) \\ &= (-)^{s-1} 2/\pi x N = (-)^{s-1} 2/\pi x \sqrt{V} \end{aligned} \quad (48)$$

by (46).

It would be possible to develop explicit expansions for M , M^{-1} , N , N^{-1} , and the right-hand sides of (47) and (48) in descending powers of x ; indeed, using the McMahon expansion, the right-hand sides of (47) and (48) might be expressed in terms of the appropriate β . The very considerable labour involved does not, at this stage, seem justified.

9. The asymptotic expansions so far given are all useful only when $x \gg n$. When x/n is less than, or not much greater than, unity, expansions of another type are necessary—and are available in the Debye and associated expansions. In these cases it is convenient to write $x = n \operatorname{sech} \alpha$ or $x = n \sec \beta^*$; we shall concern ourselves with the latter, but the corresponding formulæ for the former can readily be written down, since the same numerical coefficients are involved (W 243-4).

With the above substitution (15) yields

$$U \sim \frac{2 \cos \beta}{\pi n} \left\{ 1 + \frac{1}{2} \left(1 - \frac{1}{\mu} \right) \cos^2 \beta + \frac{1 \cdot 3}{2 \cdot 4} \left(1 - \frac{1}{\mu} \right) \left(1 - \frac{2}{\mu} \right) \cos^4 \beta + \dots \right\},$$

which after formal re-arrangement can be written

$$U \sim \frac{2}{\pi n} \left\{ u_0 - \frac{u_1}{\mu} + \frac{u_2}{\mu^2} - \frac{u_3}{\mu^3} + \dots \right\},$$

* Not to be confused with the β 's of (27) or (36).

where the u_r are power series in $\cos^2 \beta$, independent of n . In particular

$$\begin{aligned} u_0 &= \cos \beta \left\{ 1 + \frac{1}{2} \cos^2 \beta + \frac{1 \cdot 3}{2 \cdot 4} \cos^4 \beta + \dots \right\} \\ &= \cos \beta (1 - \cos^2 \beta)^{-1/2} = \cot \beta. \end{aligned}$$

Thus the leading term of U when n is large is $(2/\pi n) \cot \beta$. The series for u_1, u_2, \dots , whose coefficients involve the sums of homogeneous products of the squares of the odd integers, can be summed, but the result sought can be attained by methods less laborious.

It is easily shown, however, in a similar manner that the leading term of V is $(2/\pi n) \cos \beta (1 - \cos^2 \beta)^{1/2}$ or $(1/\pi n) \sin 2\beta$.

A little further investigation soon shows that the general results are most conveniently expressed in terms of $\cot \beta$.

10. With the notation

$$x = n \sec \beta, \quad c = \cot \beta, \quad \iota = \tan \beta, \quad \dots \quad (49)$$

so that

$$x \frac{d}{dx} = -(1 + c^2) c \frac{d}{dc}, \quad \dots \quad (50)$$

equation (13) becomes

$$\left\{ (1 + c^2) c \frac{d}{dc} \right\}^3 U = \mu \frac{1 + c^2}{c^2} \left(U - c \frac{dU}{dc} \right)$$

or, by a slight re-arrangement,

$$\frac{d}{dc} \left(\frac{U}{c} \right) = -\frac{1}{\mu} c \frac{d}{dc} \left\{ (1 + c^2) c \frac{d}{dc} \right\}^2 U. \quad \dots \quad (51)$$

In view of the results of § 9 we assume

$$U = \frac{2}{\pi n} \left\{ u_0 - \frac{u_1}{\mu} + \frac{u_2}{\mu^2} - \dots \right\}, \quad \dots \quad (52)$$

where the u_r are functions of c only. Substitution in (51) gives immediately the recurrence formula

$$\frac{d}{dc} \left(\frac{u_{r+1}}{c} \right) = c \frac{d}{dc} \left\{ (1 + c^2) c \frac{d}{dc} \right\}^2 u_r. \quad \dots \quad (53)$$

Since, by § 9,
we readily derive

$$\begin{aligned} u_1 &= c^3(1 + c^2)(1 + 5c^2)/2, & u_2 &= c^5(1 + c^2)(27 + 553c^2 + 1617c^4 + 1155c^6)/8. \\ & & & \dots \quad (54) \end{aligned}$$

If

$$u_r = \sum_{p=r}^{3r} a_{2p+1}^{(r)} c^{2p+1}, \quad \dots \quad (55)$$

(53) yields the recurrence formula

$$\begin{aligned} 2(p+1)a_{2p+3}^{(r+1)} &= (2p-3)(2p-1)(2p+1)a_{2p-3}^{(r)} \\ &\quad + 4p(2p-1)(2p+1)a_{2p-1}^{(r)} + (2p+1)^3 a_{2p+1}^{(r)}. \end{aligned} \quad (56)$$

Further coefficients u_r , calculated by (53) and checked by (56)—and also by the presence of $(1+c^2)$ as a factor—are

$$\begin{aligned} u_3 &= c^7(1+c^2)(1125+51445c^2+3\ 83570c^4+10\ 14442c^6 \\ &\quad +11\ 06105c^8+4\ 25425c^{10})/16, \\ u_5 &= c^9(1+c^2)(3\ 85875+311\ 80149c^2+4247\ 35311c^4 \\ &\quad +22221\ 56937c^6+56393\ 92473c^8+74550\ 43167c^{10} \\ &\quad +49452\ 25285c^{12}+13013\ 75075c^{14})/128, \\ u_7 &= c^{11}(1+c^2)(562\ 60575+70734\ 74499c^2+15\ 15661\ 65660c^4 \\ &\quad +128\ 47174\ 43100c^6+557\ 45928\ 61890c^8+1383\ 07094\ 91690c^{10} \\ &\quad +2045\ 57421\ 98892c^{12}+1785\ 26351\ 00300c^{14}+849\ 14723\ 64375c^{16} \\ &\quad +169\ 82944\ 72875c^{18})/256. \quad \dots \quad (54 \text{ cont.}) \end{aligned}$$

In terms of c the relation (24) becomes

$$\frac{d\psi}{dc} = -\frac{2}{\pi c(1+c^2)U}. \quad \dots \quad (55)$$

Use of the series for U in terms of c yields, on inversion,

$$\frac{d\psi}{dc} = n \left\{ -\frac{1}{c^2(1+c^2)} - \frac{1+5c^2}{2\mu} + \frac{25c^2+531c^4+1547c^6+1105c^8}{8\mu^2} - \dots \right\}. \quad (56)$$

Integrating, and determining the constant of integration with reference to (25)—noting that as $x \rightarrow \infty$, $\beta \rightarrow \frac{1}{2}\pi$ and $\tan \beta \sim x/n$ —we find

$$\begin{aligned} \psi &= n \{ t - \beta - \pi/4n - (3c+5c^3)/24n^2 \\ &\quad + (375c^3+4779c^5+9945c^7+5525c^9)/5760n^4 \\ &\quad - (67599c^5+22\ 52205c^7+130\ 77470c^9+282\ 86370c^{11} \\ &\quad +260\ 89875c^{13}+86\ 96625c^{15})/3\ 22560n^6 \\ &\quad + (56\ 35995c^7+3577\ 43715c^9+39939\ 96195c^{11} \\ &\quad +1\ 76856\ 16515c^{13}+3\ 88952\ 63673c^{15} \\ &\quad +4\ 53640\ 37985c^{17}+2\ 69226\ 62025c^{19} \\ &\quad +64101\ 57625c^{21})/34\ 40640n^8 \\ &\quad - (18276\ 97575c^9+18\ 89808\ 40233c^{11}+343\ 11310\ 89804c^{13} \\ &\quad +2521\ 59047\ 85060c^{15}+9655\ 21152\ 40482c^{17} \\ &\quad +21432\ 94995\ 27726c^{19}+28679\ 39021\ 52252c^{21} \\ &\quad +22852\ 50531\ 91380c^{23}+9999\ 87489\ 39375c^{25} \\ &\quad +1851\ 82868\ 40625c^{27})/778\ 56768n^{10} \\ &\quad + \dots \} \quad \dots \quad (57) \end{aligned}$$

This is the Meissel expansion, carried two stages further than on (W 228)* and using a different variable. It does not seem worth while throwing M into the exponential form $[e^{-Pn}$ of (W 228)].

* There is an error in Watson. The coefficient of $\sec^2 \beta$ in the term ν^{-5} should be 78720, and not 98720.

The series can evidently be used to determine the early zeros of functions of moderate or large order. For if $\mathcal{E}_n(x)=0$,

$$\psi + \alpha = (2s-1)\pi/2,$$

so that a first approximation to the zero is given by

$$t - \tan^{-1} t = \{(4s-1)\pi - 4\alpha\}/4n, \quad . \quad . \quad . \quad (58)$$

and this can be improved by successive approximation. The method works up to the accuracy of existing tables of the inverse tangent*; further work is contemplated when the American 12-decimal table becomes available.

When $\beta \rightarrow \pi/2$, (57) should allow the results of § 5 to be recovered; this has been verified as regards a few early terms.

11. On the basis of the results of § 9, V and χ can be similarly treated. In terms of c (12) becomes

$$c \frac{d}{dc} \left\{ \frac{1+c^2}{c^2} V \right\} = -\frac{1}{c} \frac{dU}{dc}. \quad . \quad . \quad . \quad (59)$$

Writing

$$V = \frac{2}{\pi n} \Sigma (-)^r \frac{v_r}{\mu_r} \quad . \quad . \quad . \quad (60)$$

and using (52), we derive

$$c \frac{d}{dc} \left\{ \frac{1+c^2}{c^2} v_r \right\} = -\frac{1}{c} \frac{du_r}{dc}. \quad . \quad . \quad . \quad (61)$$

By use of (55) we find

$$v_r = -\frac{1}{1+c^2} \Sigma_{p=r}^{3r} \frac{(2p+1)}{(2p-1)} a_{2p+1}^{(r)} c^{2p+1}. \quad . \quad . \quad . \quad (62)$$

Division of most of the numerical coefficients by $(2p-1)$ and of the expression by $(1+c^2)$ constitute an almost complete check of u_r and v_r . We find

$$\begin{aligned} v_0 &= c/(1+c^2), \quad v_1 = -c^3(3+7c^2)/2, \\ v_2 &= -c^5(45+767c^2+2023c^4+1365c^6)/8 \\ v_3 &= -c^7(1575+66015c^2+4 \ 65670c^4+11 \ 86526c^6 \\ &\quad +12 \ 60259c^8+4 \ 75475c^{10})/16 \\ v_4 &= -c^9(4 \ 96125+380 \ 84571c^2+5007 \ 24609c^4 \\ &\quad +25533 \ 81831c^6+63563 \ 74167c^8+82785 \ 83313c^{10} \\ &\quad +54269 \ 76555c^{12}+14145 \ 38125c^{14})/128 \\ v_5 &= -c^{11}(687 \ 62925+83572 \ 87617c^2+17 \ 46884 \ 51028c^4 \\ &\quad +145 \ 30996 \ 38900c^6+621 \ 31883 \ 49030c^8 \\ &\quad +1523 \ 47774 \ 10190c^{10}+2231 \ 70506 \ 31876c^{12} \\ &\quad +1932 \ 24899 \ 10724c^{14}+912 \ 91461 \ 51125c^{16} \\ &\quad +181 \ 54182 \ 29625c^{18})/256. \quad . \quad . \quad . \quad (63) \end{aligned}$$

* Comrie, 'Tracts for Computers,' xxiii.

In terms of c (33) becomes

$$c \frac{d\chi}{dc} = -\frac{2}{\pi} \frac{1}{(1+c^2)^2 V} \dots \dots \dots (64)$$

Using (63), and carrying out the inversion and integration, determining the constant as before, we find

$$\frac{d\chi}{dc} = n \left\{ -\frac{1}{c^2(1+c^2)} + \frac{3+7c^2}{8n^2} - \frac{63c^2+869c^4+2205c^6+1463c^8}{128n^4} + \dots \right\}. \quad (65)$$

$$\begin{aligned} \chi = & -\frac{1}{2}\pi + n\{t - \beta + (9c + 7c^3)/24n^2 \\ & - (945c^5 + 7821c^5 + 14175c^7 + 7315c^9)/5760n^4 \\ & + (1 \ 19637c^5 + 32 \ 40135c^7 + 173 \ 41450c^9 + 357 \ 59430c^{11} \\ & + 319 \ 39425c^{13} + 104 \ 00691c^{15})/3 \ 22560n^6 \\ & - (81 \ 52245c^7 + 4673 \ 69805c^9 + 49620 \ 85485c^{11} + 2 \ 12738 \ 37165c^{13} \\ & + 4 \ 57266 \ 43767c^{15} + 5 \ 24199 \ 06735c^{17} \\ & + 3 \ 06927 \ 95175c^{19} + 72285 \ 65015c^{21})/34 \ 40640n^8 \\ & + (23842 \ 86597c^9 + 23 \ 28192 \ 76131c^{11} + 408 \ 72316 \ 20516c^{13} \\ & + 2934 \ 71784 \ 65484c^{15} + 11044 \ 90255 \ 14630c^{17} \\ & + 24191 \ 88952 \ 62330c^{19} + 32025 \ 85744 \ 90932c^{21} \\ & + 25295 \ 33982 \ 14748c^{23} + 10987 \ 50918 \ 60765c^{25} \\ & + 2022 \ 00164 \ 89211c^{27})/778 \ 56768n^{10} - \dots\}. \quad (66) \end{aligned}$$

From this a first approximation to a zero of $\mathcal{C}'_n(x)$ analogous to (58) is

$$t - \tan^{-1} t = \{(4s-3)\pi - 4\alpha\}/4n. \quad \dots \dots \dots (67)$$

Successive approximation is again possible.

Finally, we may note the identity corresponding to (19) in terms of c

$$UV = \frac{4c^2}{\pi^2 n^2 (1+c^2)} \left\{ 1 + \pi^2 \frac{(1+c^2)^2}{16} \left(c \frac{dU}{dc} \right)^2 \right\}. \quad \dots \dots \dots (68)$$

12. The expansions of §§9-11 are closely connected with—indeed, alternative to—those of Debye (W243-4). The latter are usually obtained by the method of steepest descents, but once their existence and form are demonstrated, elementary methods prove more serviceable in developing additional terms.

In view of the known results, we write

$$H_n^{-1}(n \sec \beta) = \sqrt{\frac{2}{\pi n}} e^{in(t-\beta) - \frac{1}{2}i\pi} c^{1/2} Z(c), \quad \dots \dots \dots (69)$$

and find that Bessel's differential equation is satisfied if

$$2in \frac{dZ}{dc} = c^2(1+c^2) \frac{d^2Z}{dc^2} + 2c(1+2c^2) \frac{dZ}{dc} + \frac{1}{4}(1+5c^2)Z. \quad \dots \dots (70)$$

Writing

$$Z(c) = \sum_{r=0}^{\infty} (8in)^{-r} z_r(c). \quad \dots \dots \dots (71)$$

where $z_0=1$, it follows that

$$\frac{dz_{r+1}}{dc} = 4c^2(1+c^2) \frac{d^2z_r}{dc^2} + 8c(1+2c^2) \frac{dz_r}{dc} + (1+5c^2)z_r, \quad (72)$$

so that

$$z_0=1, \quad z_1=(3c+5c^3)/3, \quad z_2=(81c^2+462c^4+385c^6)/18, \quad (73)$$

Writing

$$z_r = \sum_{p=0}^r b_p^{(r)} c^{r+2p}, \quad (74)$$

it is found that

$$(r+2p+1)b_p^{(r+1)} = (2r+4p+1)\{(2r+4p+1)b_p^{(r)} + (2r+4p-3)b_{p-1}^{(r)}\}. \quad (75)$$

Using this recurrence formula* the successive terms of the series are readily developed.

$$z_3 = (30375c^3 + 3 \ 69603c^5 + 7 \ 65765c^7 + 4 \ 25425c^9)/810$$

$$z_4 = (44 \ 65125c^4 + 941 \ 21676c^6 + 3499 \ 22430c^8 + 4461 \ 85740c^{10} + 1859 \ 10725c^{12})/9720$$

$$z_5 = (15190 \ 35525c^5 + 4 \ 92869 \ 48607c^7 + 28 \ 44997 \ 69554c^9 + 61 \ 41358 \ 72350c^{11} + 56 \ 60981 \ 57625c^{13} + 18 \ 86993 \ 85875c^{15})/2 \ 04120$$

$$z_6 = (275 \ 70494 \ 77875c^6 + 12757 \ 72983 \ 54750c^8 + 1 \ 05076 \ 07744 \ 57901c^{10} + 3 \ 36903 \ 20682 \ 61860c^{12} + 5 \ 10469 \ 67162 \ 44125c^{14} + 3 \ 68529 \ 90061 \ 38750c^{16} + 1 \ 02369 \ 41683 \ 71875c^{18})/183 \ 70800 \dots \quad (75 \text{ cont.})$$

If we write

$$Z = Z_x + iZ_y, \quad (76)$$

then, from (21),

$$U = M^2 = \frac{2c}{\pi n} (Z_x^2 + Z_y^2), \quad (77)$$

and

$$\psi = n(t - \beta) - \frac{1}{4}\pi + \tan^{-1}(Z_y/Z_x), \quad (78)$$

and it has been verified that these relations lead to the early terms of the expressions for U and ψ given in earlier sections.

13. The Debye series for derivatives does not seem to have received much (if any) explicit attention. It can be obtained formally by differentiation, using

$$n(d/dx) = -c^2 \sqrt{(1+c^2)} \cdot (d/dc). \quad (79)$$

Writing

$$H_n^{(1)}(n \sec \beta) = \sqrt{\frac{2}{\pi n}} e^{in(t-\beta) + \frac{1}{2}i\pi} \frac{c^{1/2}}{(1+c^2)^{1/2}} W \quad (80)$$

and

$$W = \Sigma (8in)^{-r} w_r(c), \quad (81)$$

* A recurrence formula equivalent to (75) is given, without proof, in Jahnke und Emde, 'Funktionentafeln,' 3rd ed., p. 139, while in the 2nd and 3rd editions, some early z_r are given with decimal coefficients. Airey (Phil. Mag. xxxii. pp. 7-16, 1916) gives equivalent values. End figures are in both instances unreliable.

where $w_0=1$, the result of differentiating (69) leads to

$$w_r = z_r - 4c(1+c^2) \left(z_{r-1} + 2c \frac{dz_{r-1}}{dc} \right), \dots \dots \dots (82)$$

which, by use of (74) and (75), yields

$$w_r = - \sum_{p=0}^r \frac{2r+4p+1}{2r+4p-1} b_p^{(r)} c^{r+2p}, \dots \dots \dots (83)$$

so that the w_r are readily calculated. Thus

$$\begin{aligned} w_0 &= 1, & w_1 &= -(9c+7c^3)/3, & w_2 &= -(135c^2+594c^4+455c^6)/18 \\ w_3 &= -(42525c^3+4 \ 51737c^5+8 \ 83575c^7+4 \ 75475c^9)/810 \\ w_4 &= -(57 \ 40875c^4+1112 \ 34708c^6+3965 \ 78754c^8+4931 \ 52660c^{10} \\ &\quad +2020 \ 76875c^{12})/9720 \\ w_5 &= -(18565 \ 98975c^5+5 \ 68695 \ 56085c^7+31 \ 79703 \ 30678c^9 \\ &\quad +67 \ 26250 \ 03050c^{11}+61 \ 13860 \ 10235c^{13} \\ &\quad +20 \ 17131 \ 36625c^{15})/2 \ 04120 \\ w_6 &= -(325 \ 83312 \ 01125c^6+14458 \ 76048 \ 02050c^8 \\ &\quad +1 \ 16136 \ 71717 \ 69259c^{10}+3 \ 66199 \ 13785 \ 45500c^{12} \\ &\quad +5 \ 48282 \ 23989 \ 28875c^{14}+3 \ 92306 \ 02323 \ 41250c^{16} \\ &\quad +1 \ 08219 \ 09779 \ 93125c^{18})/183 \ 70800 \dots \dots \dots (84) \end{aligned}$$

There are evidently relations between V , N , χ and W analogous to those between U , M , ψ and Z given in (77) and (78).

It is also simple to derive formulæ for J , Y , \mathcal{C} , J' , Y' and \mathcal{C}' from those for H^1 and $H^{1'}$.

To my friend Dr. J. C. P. Miller my grateful thanks are once again due; his vigilance and labour with MS. and proof have resulted in the removal of error, obscurity and inconsistency, and encourage the hope that no error has escaped detection. To Professor G. N. Watson I also owe useful suggestions, which I gratefully acknowledge.

VI. Free and Forced Oscillations of Continuous Beams : Treatment by the Admittance Method.

By W. J. DUNCAN, D.Sc., M.I.Mech.E., F.R.Ae.S.,
Wakefield Professor of Aeronautics at University College, Hull*.

[Received August 7, 1942.]

SUMMARY.

The admittance method is readily applicable to the problem of determining the steady oscillation of a continuous beam of the most general kind due to impressed normal simple harmonic forces, with end

* Communicated by the Author.

thrust or tension present or absent. The method leads to a theorem of three moments, and, when impressed forces are absent, the equation of three moments is homogeneous in the amplitudes of the bending moments at the supports, while the coefficients (the angular admittances) are functions of the frequency. The eliminant of the set of such three moment equations is the frequency equation for the free oscillations of the continuous beam.

CONTENTS.

	Page
§ 1. Introduction	50
§ 2. Angular Flexibilities and Admittances of a Beam	51
§ 3. General Form of the Theorem of Three Moments for Static Loads	52
§ 4. Forced Oscillations of a Continuous Beam	53
Table I.—Values of the Angular Flexibilities for a Uniform Simply Supported Span AB	54
Table II.—Values of the Angular Admittances for a Uniform Simply Supported Span AB	55
Table III.—Static Values of the Slopes at the Supports for a Uniform Simply Supported Span AB	56
Table IV.—Amplitudes of the Oscillations in Slope at the Supports of a Uniform Simply Supported Span AB	59
§ 5. Free Oscillations of a Continuous Beam and the Calculation of Natural Frequencies	62
§ 6. Treatment of Beams with Built-in Ends	63
List of References	63

§ 1. *Introduction.*

Admittances are flexibilities measured under the conditions of steady forced oscillations of given frequency and identical phase. For any given body or system they are functions of the frequency. Their use in the calculation of the natural frequencies of systems has been explained in a quite general manner in ref. ⁽¹⁾. In the present paper the method is applied to the relatively simple problem of the forced and free oscillations of continuous beams. It is shown that, whatever the nature of the beam (*i. e.*, manner of variation of flexural rigidity and mass per unit run) and whatever may be the nature of the spanwise loads (provided that they are independent of time), there exists a theorem of three moments in which the constant term depends on the nature and distribution of the normal simple harmonic loads. When these normal loads are absent, the equation of three moments is homogeneous (and of course linear) in the amplitudes of the bending moments at the supports, and the eliminant of the whole set of these equations is the frequency equation for the continuous beam.

The problem of the oscillations of continuous beams has been discussed by Cowley and Levy ⁽²⁾. These authors obtained the theorem of three

moments in a number of important cases and showed how to obtain the natural frequencies from it. The admittance method allows the problem to be discussed in the most concise and general manner, and has the practical convenience that all the coefficients in the equations can be calculated piecemeal and at leisure; alternatively, they may be found experimentally. Tables I. to IV. of the present paper give values of the various coefficients and constants which occur in static and dynamic problems concerning continuous beams composed of uniform spans, with tensile or compressive spanwise loads present or absent. Finally, it may be emphasized that the method is applicable to beams or beam-like bodies of the most general kind.

§ 2. Angular Flexibilities and Admittances of a Beam.

Consider a single span of a beam, provided with rigid simple supports at P on the left and Q on the right. This may be an isolated span, but if it is actually part of a continuous beam it must, at present, be supposed severed from adjacent spans and treated as independent. The section of the beam may vary in any manner and it may be subject to any specified system of static transverse and end loads, which, however, must not render the span unstable.

Let a unit bending couple (positive or clockwise) be applied to the end of the beam at P, and let the angular deflexions or changes of slope due to the couple at P and Q be g_{PP} and g_{QP} respectively. Similarly, let the angular deflexions at P and Q due to the application of unit positive bending couple at Q be g_{PQ} and g_{QQ} respectively. Then, provided that the beam obeys Hooke's law, the total angular deflexions at P and Q under the action of couples N_P , N_Q are, respectively,

$$\phi_P = g_{PP}N_P + g_{PQ}N_Q \quad \dots \dots \dots (2.1)$$

$$\phi_Q = g_{QP}N_P + g_{QQ}N_Q \quad \dots \dots \dots (2.2)$$

The coefficients g are the angular flexibilities*, and by the reciprocal theorem

$$g_{PQ} = g_{QP} \quad \dots \dots \dots (2.3)$$

Also, whenever the beam is symmetrical about midspan

$$g_{PP} = g_{QQ} \quad \dots \dots \dots (2.4)$$

In general, the values of the flexibilities depend upon the load system. They are, however, independent of the transverse loads provided that these produce only small deflexions, and they are also independent of the spanwise loads if these are only small fractions of the loads which would cause the first instability. Some values for the cases where large end loads are present and absent are listed in Table I. When several spans are under consideration, they will be numbered and the cor-

* Angular influence coefficient is an alternative name.

responding number prefixed to the symbols for the flexibilities * and other quantities.

The admittances γ are defined in the same way as the flexibilities, but in relation to applied sinusoidal couples of unit amplitude and specified frequency. Accordingly, the amplitudes of the angular movements at the supports are given by

$$\phi_P = \gamma_{PP} N_P + \gamma_{PQ} N_Q, \quad \dots \dots \dots (2.5)$$

and

$$\phi_Q = \gamma_{QP} N_P + \gamma_{QQ} N_Q, \quad \dots \dots \dots (2.6)$$

where N_P and N_Q are the amplitudes of the applied periodic couples, which have the same frequency f and are in phase (or 180 deg. out of phase, which is indicated by a reversal of sign). The admittances are, in general, functions of the frequency (and of the loading, particularly the spanwise loading, just as for the flexibilities), but if the frequency is small in relation to the fundamental free frequency of the span, they become independent of frequency and equal to the corresponding static flexibilities. The admittances satisfy the reciprocal relation

$$\gamma_{PQ} = \gamma_{QP} \quad \dots \dots \dots (2.7)$$

and a relation similar to (2.4) when the beam is symmetrical about midspan. Some values of the admittances are given in Table II.

§ 3. General Form of the Theorem of Three Moments for Static Loads.

As an introduction and preliminary to the discussion of the dynamic problems of continuous beams the static theorem of three moments will be given a general formulation, since the principles are the same for the dynamic and static cases.

Consider a pair of adjacent spans of a continuous beam, say the n th PQ, and the $(n+1)$ th QR. The specification of each span must conform with that given at the beginning of § 2, but within that specification the spans with their loads may differ in any way. It is not necessary that the supports shall all be at the same level, and when they are not, the direction of the spanwise co-ordinate axis is to some extent arbitrary. For definiteness it may be taken as the straight line joining the end supports.

Imagine first that the beam is cut through at each support, so that no moment is conveyed from span to span, and calculate the slope for each span, at each of its supports, under the external load system, including of course all the spanwise loads †. In this condition there is zero bending moment at every support, accompanied, however, by discontinuities of slope. The actual bending moments at the supports are such that these discontinuities all disappear. At the support Q, for

* Without this, confusion would arise as to the meaning of g_{PP} when P is common to two spans.

† These slopes take account directly of any differences of level of the supports.

example, if the bending moment is M_Q , the clockwise couple applied to PQ at Q is

$${}_nN_Q = M_Q, \quad \dots \quad (3.1)$$

and the clockwise couple applied to QR at Q is

$$({}_{n+1})N_Q = -M_Q. \quad \dots \quad (3.2)$$

By (2.2) the changes of slope at Q due to the application of the moments are

$$\Delta_n\phi_Q = {}_ng_{QP}{}_nN_P + {}_ng_{QQ}{}_nN_Q = -{}_ng_{QP}M_P + {}_ng_{QQ}M_Q \quad \dots \quad (3.3)$$

and

$$\begin{aligned} \Delta_{(n+1)}\phi_Q &= ({}_{n+1})g_{QQ}({}_{n+1})N_Q + ({}_{n+1})g_{QR}({}_{n+1})N_R \\ &= -({}_{n+1})g_{QQ}M_Q + ({}_{n+1})g_{QR}M_R. \quad \dots \quad (3.4) \end{aligned}$$

Now let the slope of the supposedly isolated span PQ at Q be ${}_n\phi_Q$, while that of QR at Q is $({}_{n+1})\phi_Q$. Then the condition for continuity of slope at Q is

$$\Delta_{(n+1)}\phi_Q - \Delta_n\phi_Q = {}_n\phi_Q - ({}_{n+1})\phi_Q, \quad \dots \quad (3.5)$$

and on substitution from (3.3) and (3.4) the theorem of three moments is obtained in the general form

$${}_ng_{QP}M_P - ({}_ng_{QQ} + ({}_{n+1})g_{QQ})M_Q + ({}_{n+1})g_{QR}M_R = {}_n\phi_Q - ({}_{n+1})\phi_Q. \quad (3.6)$$

In this equation the expression on the right and the coefficients of the bending moments are all directly calculable. Some values of the flexibilities are given in Table I., and expressions for the end slopes ϕ for various loadings of a uniform beam are listed in Table III.

Example.

A pair of adjacent spans AB and BC, each uniform and uniformly loaded (end thrust absent).

From Tables I. and III. :

$$\begin{aligned} {}_1g_{AA} &= {}_1g_{BB} = l_1/3E_1I_1, \\ {}_1g_{AB} &= {}_1g_{BA} = -l_1/6E_1I_1, \\ {}_1\phi_A &= w_1l_1^3/24E_1I_1 = -{}_1\phi_B, \end{aligned}$$

and there are similar formulæ for the quantities appertaining to the second span. Hence equation (3.6) yields

$$\frac{l_1M_A}{E_1I_1} + \left(\frac{2l_1}{E_1I_1} + \frac{2l_2}{E_2I_2} \right) M_B + \frac{l_2M_C}{E_2I_2} = \frac{w_1l_1^3}{4E_1I_1} + \frac{w_2l_2^3}{4E_2I_2},$$

which is the familiar form of the theorem of three moments.

§ 4. Forced Oscillations of a Continuous Beam.

The system of loads considered consists of arbitrary end loads (possibly absent) which are invariable with time, and any distribution of simple harmonic transverse loads, all of which have the same frequency and are in phase (or exact opposition). Let the symbols ϕ now stand for the amplitudes of the angular movements at the supports for the individual spans, treated as independent and simply supported. Then

TABLE I.
Values of the Angular Flexibilities for a Uniform Simply Supported Span AB.

End loading.	Flexibilities.	Remarks.
Absent	$g_{AA}=g_{BB}=l/3EI$. $g_{AB}=g_{BA}=-l/6EI$. l =length of span. EI =flexural rigidity.	The special functions appearing in the middle column are known as "Berry Functions." All of them take the value 1 when a tends to zero*.
Constant thrust P	$g_{AA}=g_{BB}=\frac{1}{Pl}(1-2a \cot 2a) \mp \frac{l}{3EI} \phi(a)$. $g_{AB}=g_{BA}=-\frac{1}{Pl}(2a \operatorname{cosec} 2a-1) = -\frac{l}{6EI} f(a)$, where $a=\frac{l}{2} \sqrt{\frac{P}{EI}}$.	$\phi(a)=\frac{3}{4} \frac{(1-2a \cot 2a)}{a^2}$. $f(a)=\frac{3}{2} \frac{(2a \operatorname{cosec} 2a-1)}{a^2}$.
Constant tension T	$g_{AA}=g_{BB}=\frac{1}{Tl}(2a \coth 2a-1) = \frac{1}{3EI} \Phi(a)$. $g_{AB}=g_{BA}=-\frac{1}{Tl}(1-2a \operatorname{cosech} 2a) = -\frac{l}{6EI} F(a)$, where $a=\frac{l}{2} \sqrt{\frac{T}{EI}}$.	$\Phi(a)=\frac{3}{4} \frac{(2a \coth 2a-1)}{a^2}$. $F(a)=\frac{3}{2} \frac{(1-2a \operatorname{cosech} 2a)}{a^2}$.

* The Berry functions are tabulated in 'Aeroplane Structures,' by Pippard and Pritchard.

TABLE II.
Values of the Angular Admittances for a Uniform Simply Supported Span AB.

End loading.	Admittances.	Remarks.
Absent	$\gamma_{AA} = \gamma_{BB} = \frac{\coth al - \cot al}{2aEI} = \frac{l}{3EI} \left[\frac{3(\coth al - \cot al)}{2al} \right].$ $\gamma_{AB} = \gamma_{BA} = -\frac{\operatorname{cosec} al - \operatorname{cosech} al}{2aEI}.$ $= -\frac{l}{6EI} \left[\frac{3(\operatorname{cosec} al - \operatorname{cosech} al)}{al} \right], \text{ where } a = \sqrt[4]{\frac{m\omega^2}{EI}}.$	<p>The functions in square brackets tend to the value 1 as a tends to 0. Thus, as a tends to zero, the expressions for the admittances tend to equality with the corresponding flexibilities (see Table I). m = mass per unit length. $\omega = 2\pi$ times frequency.</p>
Constant thrust P . . .	$\gamma_{AA} = \gamma_{BB} = \frac{(a \coth al - \beta \cot \beta l)}{EI(a^2 + \beta^2)}.$ $\gamma_{AB} = \gamma_{BA} = -\frac{(\beta \operatorname{cosec} \beta l - a \operatorname{cosech} al)}{EI(a^2 + \beta^2)}.$ <p>where $a^2 = \frac{1}{2}(\sqrt{4a^4 + \mu^4} - \mu^2)$, $\beta^2 = \frac{1}{2}(\sqrt{4a^4 + \mu^4} + \mu^2)$.</p> $a = \sqrt[4]{\frac{m\omega^2}{EI}}, \quad \mu = \sqrt{\frac{P}{EI}}.$	<p>When P is zero, μ vanishes, and $a = \beta = a$. The expressions for the admittances then reduce to those given above. <i>Note</i> that a has not the same meaning here as in Table I. When a tends to zero while μ is finite, $a \rightarrow 0$ and $\beta \rightarrow \mu$. The admittances then become equal to the flexibilities given in Table I. for the case of end thrust.</p>
Constant tension T . .	$\gamma_{AA} = \gamma_{BB} = \frac{(\beta \coth \beta l - a \cot al)}{EI(a^2 + \beta^2)}.$ $\gamma_{AB} = \gamma_{BA} = -\frac{(a \operatorname{cosec} al - \beta \operatorname{cosech} \beta l)}{EI(a^2 + \beta^2)}.$ <p>where α and β are as defined immediately above and $\mu = \sqrt{\frac{T}{EI}}$.</p>	<p>The remarks are essentially the same as for the case of end thrust P (above).</p>

TABLE III.
Static Values of the Slopes at the Supports for a Uniform Simply Supported Span AB.

End loading.	Transverse loading.	Slope.	Remarks.
Absent	Isolated load W at midspan.	$\phi_A = -\phi_B = \frac{Wl^2}{16EI}$.	
	Isolated load W at distance h from A.	$\phi_A = \frac{Wh(l-h)(2l-h)}{6EI}$, $\phi_B = -\frac{Wh(l-h)(l+h)}{6EI}$.	
	Uniformly distributed, w per unit run.	$\phi_A = -\phi_B = \frac{wl^3}{24EI}$.	
	Variable distributed, $w(x)$ per unit run.	$\phi_A = -\frac{1}{lEI} \int_0^l (l-x)M(x) dx$, $= \frac{1}{6EI} \int_0^l w(x)x(l-x)(2l-x) dx$, $\phi_B = \frac{1}{lEI} \int_0^l xM(x) dx = -\frac{1}{6EI} \int_0^l w(x)x(l-x)(l+x) dx$.	x is measured from the support A. $M(x)$ is the bending moment at distance x from A.
Constant thrust P.	Isolated load W at midspan.	$\phi_A = -\phi_B = \frac{W}{2P} (\sec \alpha - 1) = \frac{Wl^2}{16EI} \left[\frac{2(\sec \alpha - 1)}{\alpha^2} \right]$.	$\alpha = \frac{l}{2} \sqrt{\frac{P}{EI}}$. The function in square brackets tends to unity when α tends to zero.

TABLE III. (*cont.*).

End loading.	Transverse loading.	Slope.	Remarks.
Constant thrust P.	Isolated load W at distance h from A.	$\phi_A = \frac{W}{P} \left\{ \frac{\sin \mu(l-h)}{\sin \mu l} - \frac{l-h}{l} \right\}$ $= \frac{Wh(l-h)(2l-h)\psi(l-h, l, \mu)}{6lEI}$ $\phi_B = -\frac{W}{P} \left\{ \frac{\sin \mu h}{\sin \mu l} - \frac{h}{l} \right\}$ $= -\frac{Wh(l-h)(l+h)\psi(h, l, \mu)}{6lEI}$	$\mu = \sqrt{\frac{2a}{EI}}$ $\psi(x, l, \mu) = \frac{\mu^2 x(l^2 - x^2) \sin \mu l}{6(l \sin \mu x - x \sin \mu l)}$ <p>This function tends to 1 as μ tends to 0.</p>
	Uniformly distributed, w per unit run.	$\phi_A = -\phi_B = \frac{wl}{2P} \left(\frac{\tan \alpha}{\alpha} - 1 \right)$ $= \frac{wl^3}{24EI} \psi(\alpha).$	$\psi(\alpha)$ is the "Berry Function" $\frac{3(\tan \alpha - \alpha)}{\alpha^3}$. This tends to 1 as α tends to 0.
	Variable distributed, $w(x)$ per unit run.	$\phi_A = \frac{1}{P} \int_0^l w(x) \left\{ \frac{\sin \mu(l-x)}{\sin \mu l} - \frac{l-x}{l} \right\} dx$ $= \frac{1}{6lEI} \int_0^l w(x)x(l-x)(2l-x)\psi(l-x, l, \mu) dx.$ $\phi_B = -\frac{1}{P} \int_0^l w(x) \left\{ \frac{\sin \mu x}{\sin \mu l} - \frac{x}{l} \right\} dx.$ $= -\frac{1}{6lEI} \int_0^l w(x)x(l-x)(l+x)\psi(x, l, \mu) dx.$	

TABLE III. (cont.).

End loading.	Transverse loading.	Slope.	Remarks.
Constant tension T.	Isolated load W at midspan.	$\phi_A = -\phi_B = -\frac{W}{2T}(1 - \operatorname{sech} a) = \frac{Wl^2}{16EI} \left[\frac{2(1 - \operatorname{sech} a)}{a^2} \right]$	$a = \frac{l}{2} \sqrt{\frac{T}{EI}}$. The function in square brackets tends to 1 when a tends to 0.
	Isolated load W at distance h from A.	$\phi_A = \frac{W}{T} \left\{ \frac{l-h}{l} - \frac{\sinh \mu(l-h)}{\sinh \mu l} \right\} = \frac{Wh(l-h)(2l-h)\Psi(l-h, l, \mu)}{6EI}$ $\phi_B = -\frac{W}{T} \left\{ \frac{h}{l} - \frac{\sinh \mu h}{\sinh \mu l} \right\} = -\frac{Wh(l-h)(l+h)\Psi(h, l, \mu)}{6EI}$	$\mu = \sqrt{\frac{T}{EI}}$, $\frac{2a}{l}$ $\Psi(x, l, \mu) = \frac{\mu^2 x \sinh \mu l - l \sinh \mu x}{6(x \sinh(l^2 - x^2) \sinh \mu l)}$. This function tends to 1 as μ tends to 0.
	Uniformly distributed, w per unit run.	$\phi_A = -\phi_B = \frac{wl}{2T} \left(1 - \frac{\tanh a}{a} \right) = \frac{wl^3}{24EI} \Psi(a)$.	$\Psi(a)$ is the "Berry Function", $\frac{3(a - \tanh a)}{a^3}$.
	Variable distributed, $w(x)$ per unit run.	$\phi_A = \frac{1}{T} \int_0^l w(x) \left\{ \frac{l-x}{l} - \frac{\sinh \mu(l-x)}{\sinh \mu l} \right\} dx$ $= \frac{1}{6EI} \int_0^l w(x)x(l-x)(2l-x)\Psi(l-x, l, \mu) dx$ $\phi_B = -\frac{1}{T} \int_0^l w(x) \left\{ \frac{x}{l} - \frac{\sinh \mu x}{\sinh \mu l} \right\} dx$ $= -\frac{1}{6EI} \int_0^l w(x)x(l-x)(l+x)\Psi(x, l, \mu) dx$.	

TABLE IV.
Amplitudes of the Oscillations in Slope at the Supports of a Uniform Simply Supported Span AB.
(All loads simple harmonic, of same frequency, and in phase.)

End loading.	Transverse loading.	Slope.	Remarks.
Absent	Isolated load W at midspan.	$\phi_A = -\phi_B = \frac{W}{4a^2EI} \left(\sec \frac{al}{2} - \operatorname{sech} \frac{al}{2} \right) = \frac{Wl^2}{16EI} \left[\frac{4 \left(\sec \frac{al}{2} - \operatorname{sech} \frac{al}{2} \right)}{a^2 l^2} \right],$ <p>where</p> $a = \sqrt[4]{\frac{\omega^2}{EI}}.$	The function in square brackets tends to 1 as a tends to 0.
	Isolated load W at distance h from A.	$\phi_A = \frac{W}{2a^2EI} \left(\frac{\sin a(l-h)}{\sin al} - \frac{\sinh a(l-h)}{\sinh al} \right).$ $\phi_B = -\frac{W}{2a^2EI} \left(\frac{\sin ah}{\sin al} - \frac{\sinh ah}{\sinh al} \right).$	
	Uniformly distributed, w per unit run.	$\phi_A = -\phi_B = \frac{w}{2a^3EI} \left(\tan \frac{al}{2} - \tanh \frac{al}{2} \right)$ $- \frac{wl^3}{24EI} \left[\frac{12 \left(\tan \frac{al}{2} - \tanh \frac{al}{2} \right)}{a^3 l^3} \right].$	As above.
	Variable distributed, $w(x)$ per unit run.	$\phi_A = \frac{1}{2a^2EI} \int_0^l w(x) \left\{ \frac{\sin a(l-x)}{\sin al} - \frac{\sinh a(l-x)}{\sinh al} \right\} dx.$ $\phi_B = -\frac{1}{2a^2EI} \int_0^l w(x) \left\{ \frac{\sin ax}{\sin al} - \frac{\sinh ax}{\sinh al} \right\} dx.$	

TABLE IV. (cont.).

End loading.	Transverse loading.	Slope.	Remarks.
Constant thrust P.	Isolated load W at midspan.	$\phi_A = -\phi_B = \frac{W}{2EI(\alpha^2 + \beta^2)} \left(\sec \frac{\beta l}{2} - \operatorname{sech} \frac{\alpha l}{2} \right).$ $\alpha^2 = \frac{1}{2}(\sqrt{4\alpha^4 + \mu^4} - \mu^2), \quad \beta^2 = \frac{1}{2}(\sqrt{4\alpha^4 + \mu^4} + \mu^2).$	$\alpha = \sqrt[4]{\frac{m\omega^2}{EI}}.$ $\mu = \sqrt{\frac{P}{EI}}.$
	Isolated load W at distance h from A.	$\phi_A = \frac{W}{EI(\alpha^2 + \beta^2)} \left[\frac{\sin \beta(l-h)}{\sin \beta l} - \frac{\sinh \alpha(l-h)}{\sinh \alpha l} \right].$ $\phi_B = \frac{-W}{EI(\alpha^2 + \beta^2)} \left[\frac{\sin \beta h}{\sin \beta l} - \frac{\sinh \alpha h}{\sinh \alpha l} \right].$	
	Uniformly distributed, w per unit run.	$\phi_A = -\phi_B = \frac{w}{EI(\alpha^2 + \beta^2)} \left\{ \tan \frac{\beta l}{2} - \frac{\tanh \frac{\alpha l}{2}}{\beta} \right\}.$	
	Variable distributed, $w(x)$ per unit run.	$\phi = \frac{1}{EI(\alpha^2 + \beta^2)} \int_0^l w(x) \left\{ \frac{\sin \beta(l-x)}{\sin \beta l} - \frac{\sinh \alpha(l-x)}{\sinh \alpha l} \right\} dx.$ $\phi_B = \frac{-1}{EI(\alpha^2 + \beta^2)} \int_0^l w(x) \left\{ \frac{\sin \beta x}{\sin \beta l} - \frac{\sinh \alpha x}{\sinh \alpha l} \right\} dx.$	

TABLE IV. (cont.).

End loading.	Transverse loading.	Slope.	Remarks.
Constant tension T.	Isolated load W at midspan.	$\phi_A = -\phi = \frac{W}{2EI(\alpha^2 + \beta^2)} \left(\sec \frac{\alpha l}{2} - \operatorname{sech} \frac{\beta l}{2} \right).$ <p>α and β are as defined above.</p> $\mu = \sqrt{\frac{T}{EI}}.$	Derived from thrust case by interchange of α and β .
	Isolated load W at distance h from A.	$\phi_A = \frac{W}{EI(\alpha^2 + \beta^2)} \left[\frac{\sin \alpha(l-h)}{\sin \alpha l} - \frac{\sinh \beta(l-h)}{\sinh \beta l} \right].$ $\phi_B = \frac{-W}{EI(\alpha^2 + \beta^2)} \left[\frac{\sin \alpha h}{\sin \alpha l} - \frac{\sinh \beta h}{\sinh \beta l} \right].$	
	Uniformly distributed, w per unit run.	$\phi_A = -\phi_B = \frac{w}{EI(\alpha^2 + \beta^2)} \left(\tan \frac{\alpha l}{2} - \frac{\tanh \frac{\beta l}{2}}{\beta} \right).$	
	Variable distributed, $w(x)$ per unit run.	$\phi_A = \frac{1}{EI(\alpha^2 + \beta^2)} \int_0^l w(x) \left\{ \frac{\sin \alpha(l-x)}{\sin \alpha l} - \frac{\sinh \beta(l-x)}{\sinh \beta l} \right\} dx.$ $\phi_B = -\frac{1}{EI(\alpha^2 + \beta^2)} \int_0^l w(x) \left\{ \frac{\sin \alpha x}{\sin \alpha l} - \frac{\sinh \beta x}{\sinh \beta l} \right\} dx.$	

a theorem of three moments can be deduced in exactly the same manner as in the static case, and it is expressed by

$${}_n\gamma_{QP}M_P - ({}_n\gamma_{QQ} + {}_{(n+1)}\gamma_{QQ})M_Q + {}_{(n+1)}\gamma_{QR}M_R = {}_n\phi_Q - {}_{(n+1)}\phi_Q. \quad (4.1)$$

When the amplitudes of the bending moments have been found from the set of equations (4.1), the complete solution for the steady forced motion of each span can be found. Values of the slopes ϕ for several cases of loading of a uniform beam are given in Table IV.

Example.

A pair of adjacent spans AB and BC, both uniform, free from end load and simply supported at A and C. A uniformly distributed simple harmonic load is applied to AB.

From Tables II and IV.:

$$\begin{aligned} {}_1\gamma_{BB} &= \frac{\coth a_1 l_1 - \cot a_1 l_1}{2a_1 E_1 I_1}, \\ {}_2\gamma_{BB} &= \frac{\coth a_2 l_2 - \cot a_2 l_2}{2a_2 E_2 I_2}, \\ {}_1\phi_B &= -\frac{w_1}{2a_1^3 E_1 I_1} \left(\tan \frac{a_1 l_1}{2} - \tanh \frac{a_1 l_1}{2} \right). \end{aligned}$$

Hence by (4.1)

$$M_B = \frac{\frac{w_1}{a_1^3 E_1 I_1} \left(\tan \frac{a_1 l_1}{2} - \tanh \frac{a_1 l_1}{2} \right)}{\frac{\coth a_1 l_1 - \cot a_1 l_1}{a_1 E_1 I_1} + \frac{\coth a_2 l_2 - \cot a_2 l_2}{a_2 E_2 I_2}}.$$

Since M_B is known, the complete solutions for the two spans can be obtained.

§ 5. Free Oscillations of a Continuous Beam and the Calculation of Natural Frequencies.

When the oscillations of the beam are free the terms ϕ on the right-hand side of the equation (4.1) are absent, so that the amplitudes of the bending moments satisfy a set of linear homogeneous equations, whose coefficients (the admittances) are functions of the frequency. The eliminant or condition of compatibility of these equations provides the equation determining the natural frequencies. The roots are all real and may be found by arithmetical or graphical methods.

Example.

Any beam ABCD of three spans having the ends A, D simply supported. The equations for the bending moments are

$$\begin{aligned} -({}_1\gamma_{BB} + {}_2\gamma_{BB})M_B + {}_2\gamma_{BC}M_C &= 0, \\ {}_2\gamma_{CB}M_B - ({}_2\gamma_{CC} + {}_3\gamma_{CC})M_C &= 0. \end{aligned}$$

Hence the frequency equation is

$$({}_1\gamma_{BB} + {}_2\gamma_{BB})({}_2\gamma_{CC} + {}_3\gamma_{CC}) - {}_2\gamma_{BC}^2 = 0,$$

since

$${}_2\gamma_{BC} = {}_2\gamma_{CB}.$$

§ 6. *Treatment of Beams with Built-in Ends.*

Whether the problem under consideration is static or dynamic, the equation connecting the bending moments at the rigidly built-in end and at the adjacent support can be obtained at once from the theorem of three moments by imagining the rigid fixing replaced by a very short unloaded span provided with simple supports. When the length of this imaginary span is made to approach zero the equation appropriate to the built-in end is obtained.

Let the span be AB, built-in at A. Imagine this extended by the short span OA of length l_0 . The equation corresponding to (4.1) is

$${}_0\gamma_{AO}M_O - ({}_0\gamma_{AA} + {}_1\gamma_{AA})M_A + {}_1\gamma_{AB}M_B = {}_0\phi_A - {}_1\phi_A. \quad (6.1)$$

But as $l_0 \rightarrow 0$, ${}_0\gamma_{AO}$, ${}_0\gamma_{AA}$ and ${}_0\phi_A$ all tend to zero. Hence the equation appropriate to the built-in span is

$${}_1\gamma_{AA}M_A + {}_1\gamma_{AB}M_B = -{}_1\phi_A, \quad (6.2)$$

where ${}_1\phi_A$ is, as usual, calculated for a simply supported span under the actual loads applied to AB. Equation (6.2) could have been written down at once as the condition that the total angular deflexion at A vanishes.

References.

- (1) W. J. Duncan, "The Admittance Method for Obtaining the Natural Frequencies of Systems," *Phil. Mag.* ser. 7, vol. xxxii. p. 401 (Nov., 1941).
- (2) W. L. Cowley and H. Levy, "Vibration and Strength of Struts and Continuous Beams under End Thrust," *Proc. Roy. Soc. A*, vol. xev. p. 440 (1919).

VII. *The Separation of Electricity in Clouds.*

By J. ALAN CHALMERS, M.A., Ph.D.,

Lecturer in Physics in the Durham Colleges of the University of Durham *.

[Received November 2, 1942.]

1. *Introduction.*

SIMPSON AND SCRASE⁽¹⁾ have put forward the theory that the main process of separation of charge within clouds is one of friction between ice particles of different sizes, the larger particles acquiring negative and the smaller positive charges. Such a process may well be imagined to occur in the upper portion of a thundercloud, but not so easily in the less violent conditions of other types of rain cloud, *e. g.* those associated with

* Communicated by the Author.

a warm front, where electrical effects are also found [Chalmers and Little ⁽²⁾, etc.].

It is the purpose of the present paper to point out certain features which do not seem very satisfactory in connexion with the theory and to suggest an alternative mechanism for the separation of charge in clouds.

2. The Constituents of a Cloud.

The electrical processes within a cloud depend on the presence of water drops or ice particles, and so, before any theory of the electrical processes can be discussed, it is necessary to know the constitution of the cloud at different levels, so as to know at what levels we can assume processes depending on the presence of ice particles or water drops. Simpson and Scrase ⁽¹⁾ have, apparently, assumed that there are no water drops at temperatures below the freezing-point, but only ice particles. This is not, however, the case, as can be seen from the phenomenon of ice-accretion on aircraft flying through clouds at temperatures below the freezing-point, the type of ice being that due to the freezing on the aircraft of water droplets, analogous to the formation of glazed frost; this must mean that supercooled water droplets exist at temperatures well below the freezing-point.

Bergeron ⁽³⁾ has put forward a theory of the formation of rain, according to which rain can occur only when water drops and ice particles are present simultaneously at the same level in the cloud; then, owing to the difference in vapour pressure, the ice particles grow at the expense of the water drops and so are able to fall relatively to the drops; meeting other drops while falling, they coalesce with them and freeze them, increasing in size; finally, these ice particles melt after reaching the freezing-point level. An essential for this process is the formation of at least a few ice particles, at a suitable height, by the freezing of some supercooled water drops; it is probable that suitable dust nuclei or fragments of ice would be needed to start such freezing. It is clear that, as the temperature falls, the chance of such freezing increases and, at a low enough temperature, all the constituents of a cloud will be ice particles which cannot grow at each other's expense and so there will be no rain. This theory explains why there is no rain from, on the one hand, cirrus clouds consisting entirely of ice particles, nor, on the other hand, from stratus clouds or fog, consisting entirely of water drops.

In any cloud which contains both ice particles and water drops, there will be a region of transition in which both are present together and in which the formation of rain originates. The temperature at which the transition occurs and the depth of the region of transition must depend on a number of factors which will include the vertical air current in the cloud, the lapse rate, and the presence of dust nuclei. It will not be possible to specify any one temperature at which the transition occurs, and this must differ from one cloud to another, and possibly even in different parts of the same cloud.

3. *Transition Temperatures.*

In any individual cloud, or part of a cloud, there is a temperature, which we shall call "F," which is the temperature around which the supercooled water drops and the ice particles co-exist, and so at which the process of rain formation occurs. In some cases "F" may represent a very narrow range of temperatures and in others a broader range.

Severe ice-accretion, of the glazed frost type, can occur on aircraft in clouds of the nimbo-stratus type at temperatures down to -15°C . or possibly lower; such ice-accretion must be due to supercooled water drops and proves the presence of the drops, so that "F" is -15°C . or lower in these cases.

In strato-cumulus clouds, where there is seldom any rain, there can be no ice particles, but only water drops, and these clouds go to temperatures at least as low as -10°C ., so that "F" would be below this value.

In the more violent upward currents of thunderstorms, it seems probable, on account of the vertical air current, that the supercooled water drops might go to greater heights, and hence to lower temperatures, than in the less violent types of cloud, so we might expect the temperature "F" to be lower than -15°C . and perhaps lower than -20°C .

4. *Thundercloud Results.*

The observations by the "alti-electrograph" of Simpson and Scrase⁽¹⁾ and Simpson and Robinson⁽⁴⁾ have shown that the negative charge in a thundercloud is centred round the 0°C . isothermal, while the upper positive charge occurs at temperatures around -20°C . If the separation of charge is the process of ice-friction, as suggested by Simpson and Scrase, we might expect it only to occur at the levels where the ice particles are predominant and therefore at levels where the temperature is below "F." Since Bergeron's theory indicates that the only ice particles which will grow enough to fall in the cloud are those which have grown at the expense of the water drops at temperatures near "F," it is difficult to see how a negative charge produced by ice friction at temperatures below "F" could fall to give an accumulation of negative charge near the 0°C . level.

The results of Simpson and Scrase⁽¹⁾ and Simpson and Robinson⁽⁴⁾ suggest that the separation of charge occurs at levels round about -10°C . and, unless "F" is a higher temperature than this, contrary to the conclusions of the last section, it seems that we should search for the mechanism of separation in some process existing at levels round the temperature of -10°C ., rather than in the process of ice friction which could only occur at temperatures below "F."

5. *Non-Raining Clouds.*

Wilson⁽⁵⁾ has pointed out that non-raining clouds produce very little effect on the atmospheric potential gradient, and therefore, presumably,

have little electrical separation. However, when rain begins, there is always a definite effect, and this would seem to indicate that the electrical effects are associated with the beginning of the rain. If we compare the conditions of a nimbo-stratus cloud with those of a strato-cumulus cloud with cirrus above, we see that, according to Bergeron's theory, the rain is due to the presence of the water drops and ice particles together in the one case and not in the other. It seems a legitimate assumption that the electrical effects can be traced to the same cause, as has already been indicated by the alti-electrograph results.

6. Electrical Effects at Water Surfaces.

The experiments of Simpson ⁽⁶⁾, Nolan ⁽⁷⁾, and others have shown that the sudden production of a new air-water surface gives rise to a separation of electric charge, the water receiving a positive charge and the air a negative charge. Simpson ⁽⁶⁾ has considered the breaking of drops to be responsible for the electrification of thunderclouds, but this would give the polarity opposite to that found by the alti-electrograph for the main body of the thundercloud; however, Simpson and Scrase ⁽¹⁾ have retained the breaking-drop theory to account for the localized positive charge in the base of the cloud.

The process of the separation of charge on the production of a new air-water surface can be related to the existence of a potential difference at an air-water surface, as shown by the experiments of Chalmers and Pasquill ⁽⁸⁾. The potential difference can be ascribed to the orientation of the water molecules in the surface layer, there being a negative charge outwards and a positive charge inwards. Thus the production of a new surface must involve the negative ends of the molecules being projected outwards and, if the process is sufficiently violent, the molecules may be ruptured, the negative charge escaping into the air while the positive charge is retained in the water. Such a mechanism would explain why the separation of charge occurs only on a violent breaking of liquid, not on a mere extension of a surface.

In order to account for the separation of charge which occurs in a thundercloud around temperatures of $-10^{\circ}\text{C}.$, it is necessary to have some process in which molecules can be ruptured with the positive charge escaping into the air and the negative charge remaining. By analogy with the process just discussed, this could occur if the molecules are suddenly orientated with the positive ends outwards.

When a supercooled water drop suddenly freezes, the solidified ice particle will have its molecules arranged in a regular fashion; some of these molecules will have their negative ends outwards, as in the liquid, but some will have their positive ends outwards and so, if the freezing is sufficiently sudden, we have the condition required for positive charge to escape, leaving the negative charge behind.

A separation of charge due to sudden freezing would occur not only at the transition level, where some water drops freeze to give the first

ice-particles, but also at lower levels, where the falling ice particles coalesce with water drops, freezing the latter. So such a separation of charge will occur at all temperatures from 0°C. down to the transition temperature "F."

7. Conclusion.

The theory of the formation of rain suggests that sudden freezing of water drops occurs at temperatures from 0°C. down to the temperature "F," which in thunderclouds will be -15°C. to -20°C. The investigations with the alti-electrograph have shown that there is a separation of charge at about the same temperatures, and it seems natural to associate the two phenomena occurring in the same part of the cloud.

The present discussion has suggested how a separation of charge might be expected on sudden freezing by a mechanism similar to that which can account for the separation of charge on the breaking of a drop.

It is hoped that, when conditions permit, it may be possible to test the theory by measuring any charging of drops which freeze suddenly in a laboratory experiment. It would seem to be better to look for the effect in falling drops, where any separated charge must go into the air, rather than in the case of supercooled water in bulk suddenly freezing, where the separated charge might go to the containing vessel and where the air is not so easily removed from the surface.

References.

- (1) Simpson and Scrase, Proc. Roy. Soc. A, clxi. p. 309 (1939).
- (2) Chalmers and Little, Terr. Mag. xlv. p. 451 (1940).
- (3) Bergeron, *Union Géod. et Géophys. Intern. (Lisbon)*, p. 156 (1933).
- (4) Simpson and Robinson, Proc. Roy. Soc. A, clxxvii. p. 281 (1941).
- (5) Wilson, Dict. of App. Phys. iii. p. 84 (1923).
- (6) Simpson, Phil. Trans. Roy. Soc. A, ccix. p. 379 (1909).
- (7) Nolan, Proc. Roy. Soc. A, xc. p. 531 (1914).
- (8) Chalmers and Pasquill, Phil. Mag. xxiii. p. 88 (1937).

VIII. On Solidified Water-films.

By H. Löwy, Dr.Phil.*

[Received August 4, 1939.]

THE knowledge of the electrodynamic volume permits, as I have shown †, the measurement of the thickness of electrically conducting films covering dielectric particles which are dispersed in a dielectric medium.

An example of such a system is seen in the desert soil, which normally contains no gravitational and no capillary water, and whose moisture

* Communicated by the Author.

† H. Löwy, Phil. Mag. xxxiii. p. 772 (1942).

d_f being the mean film-density, $d_w=1$ being the normal water-density. Eliminating v_f from (2) and (3), we obtain :

$$\frac{[\epsilon_f]}{d_f} = \frac{[\bar{\epsilon}] - [\epsilon_e]v_e}{v_w} \dots \dots \dots (4)$$

If we know ϵ_f as a function of d_f , we can also exhibit $\frac{[\epsilon_f]}{d_f}$ as a function of d_f . This is done in the figure, utilizing for the water-density d_f the values which P. W. Bridgman* has measured up to pressures of 12,000 kg./cm.², and for the dielectric constant ϵ_f of water the values

TABLE I.

Pressure (kg./cm. ²).	d_f (acc. Bridgman).	ϵ_f (acc. Kyropoulos).	$\frac{[\epsilon_f]}{d_f}$
1	0.998	80.79	0.968
500	1.019	83.07	0.946
1000	1.038	85.20	0.930
1500	1.056	87.03	0.915
2000	1.072	88.72	0.900
2500	1.087	90.34	0.883
3000	1.101	91.90	0.876

TABLE II.

Pressure (kg./cm. ²).	d (acc. Bridgman).	$\frac{0.965}{d_f}$
4000	1.12	0.86
5000	1.15	0.84
6000	1.17	0.82
7000	1.19	0.81
8000	1.21	0.79 ₅
9000	1.23	0.78 ₅

measured by S. Kyropoulos† (in Jammann's laboratory) up to 3000 kg./cm.² The values in the tables and in the figure are valid for the temperature 20° C.

Measuring $\bar{\epsilon}$, ϵ_e , v_e , and v_w , we determine a certain value of the ordinate $\frac{[\epsilon_f]}{d_f}$ of the curve. The corresponding value of the abscissa is the mean density of the water-film. We have thus found a simple method for measuring the mean density, the volume, and thickness of solidified water-films.

* P. W. Bridgman, Proc. Amer. Acad. xlviii. p. 338 (1912).

† S. Kyropoulos, Zeits. f. Phys. xl. p. 507 (1927).

I intend to test this method later. The desert experiments I have so far made cannot be used for this purpose because I have not measured the quantities ϵ_c and v_w . Nevertheless, it is possible, by making plausible suppositions, to see that the method will prove useful.

We know, from my previous paper * that on March 7, 1939, in the region of the Western Desert of Egypt, where I have worked, the water-films had, accordingly to (1), a conducting surface layer. We suppose that after three rainless days, on March 10, the water-films on the soil particles have become purely dielectric substances. On this day we have measured $\bar{\epsilon}=2.1$; therefore $[\bar{\epsilon}]=0.268$ and $v_c=0.50$. The value of ϵ_c is unknown, but we know that the Mossotti function of this argument is included between the values

$$[\bar{\epsilon}] < [\epsilon_c] < \frac{[\bar{\epsilon}]}{v_c};$$

for $[\epsilon_c]$ is greater than $[\bar{\epsilon}]$ if there are no conducting particles in the rock †. The second relation follows from the fact that all quantities in equation (2) are positive. According to (5) we make the hypothesis $[\epsilon_c]=0.52$ or $\epsilon_c=4.2$. We obtain thus $[\bar{\epsilon}]-[\epsilon_c]v_c=0.008$. For the water content we assume $v_w=0.009$, a value among the smallest values which A. H. Montasir ‡ has measured in this part of the Western Desert. Thus we find :

$$\text{the film density } d_f = 1.08,$$

$$\text{the film volume } v_f = \frac{v_w}{d_f} = 8.3 \cdot 10^{-3},$$

$$\text{the film thickness } b_f = \frac{v_f}{4\pi \cdot 10^{-6} \cdot 10^6} = 6.6 \cdot 10^{-6}.$$

According to the measurements of Kyropoulos, the value $[\epsilon_f]$ is constant over the whole range of pressures from 1 to 3000 kg./cm.² I have used this value for extrapolation up to the pressures used by Bridgman in his density measurements. The extrapolated part of the curve is dashed, but has not been used in these computations.

Cairo, Egypt,
July 30, 1939.

* H. Löwy, *Phil. Mag. loc. cit.*

† H. Löwy, *Phil. Mag.* xxvii. p. 576 (1939).

‡ A. H. Montasir, "Egyptian Soil Structure in Relation to Plants" ('Bulletin of the Faculty of Science,' The Egyptian University, Cairo, 1938).

IX. *Note on a Passage in Stukeley's 'Memoirs of Sir Isaac Newton's Life.'*

IN Stukeley's *Memoirs* the following passage occurs: "I have heard it, as a tradition, whilst I was student at Cambridge, that when Sir Isaac stood for Bachelor of Arts degree, he was put to second posing, or lost his groats, as they term it; which is look'd upon as disgraceful. I can't tell whether it be true or not; but it seems no strange thing at that time of day, notwithstanding Sir Isaac's great parts; for he was too busy in the solid track of learning and the sublime pursuits of mathematical philosophy, to allow of time enough to be master of words only, or the trifling nicetys of technical logic and school subtleties, which then was the chief test of proficiency in academic learning and qualification for a degree."

Stukeley was by some forty years Newton's junior; in 1705, when Newton was knighted at Cambridge by Queen Anne, Stukeley was a young man in residence "in what we call there junior sophs or 3^d year after admission, residing in C.C. College." He became personally acquainted with Newton in London about the year 1718, and they remained on terms of familiar friendship until Newton's death in March 1727. The tradition which Stukeley mentions was, then, current in 1705, forty years after Newton's graduation.

In a paper in the December (1942) number of the *Philosophical Magazine*, dealing with Newton and the 'Principia,' I drew attention to the phrase "lost his groats" in a footnote which runs:—"The editor of Stukeley's 'Memoirs of Sir Isaac Newton's Life' (p. 53), remarks concerning the strange phrase 'lost his groats'—" 'groats,' not 'greats.' This curious expression seems quite lost, unrecorded, and unexplained. Perhaps some reader of the *Philosophical Magazine* can throw light on its origin."

Mr. F. P. White, M.A., Fellow of St. John's College, Cambridge, has kindly sent the following note, which completely clears up the obscurity:—

"The phrase 'to save his groats' is given in 'Gradus ad Cantabrigiam' (London, 1824), p. 55, with the following quotation from Grose's Dictionary of the Vulgar Tongue: 'At the Universities, nine groats are deposited in the hands of an academic officer by every person standing for a degree, which, if the depositor obtains with honour, are returned to him.'

"This may throw some light on your query in the *Philosophical Magazine*. There are various references to the payment of groats in early accounts of ceremonies of the University; the most convenient reference is probably Wordsworth: 'Social Life at the English Universities in the 18th century' (Cambridge, 1874), pp. 248, 249, 253. And the Elizabethan Statutes of 1570 require the payment of the sum of 4*d.* to various officials on admission to the B.A. (see Heywood and Wright, 'Cambridge University Transactions during the Puritan Controversies,' London, 1854, vol. i. p. 36)."

ALLAN FERGUSON.

X. Notices respecting New Books.

The Mathematical Tables Project (Federal Works Agency, New York.)

SOME work is undertaken primarily for the sheer interest of it, some for personal renown, some for the monetary reward which it will bring, whilst in some cases the dominating motive is a desire to help, and to be of service to, others. We can have little doubt that those who initiated the mathematical tables project were actuated by the motive of service, and those who are engaged in the actual computations would only have undertaken the task for the same reasons, mingled with a real enthusiasm for this aspect of mathematics.

At any rate, whatever the motives, the result is clear: a benefit conferred on mathematicians, statisticians and physicists, which they will be anxious to acknowledge, and which may in some degree lead them to provide the third in our list of motives, by purchasing copies. The labours of the mathematicians engaged in the work have already resulted in the publication of extensive tables of the exponential function $e^{\pm x}$, the sine and cosine up to 100 radians, hyperbolic functions, Planck's function, the error function and its integrand $\frac{2}{\sqrt{\pi}}e^{-x^2}$, the sine, cosine and exponential integrals, natural logarithms and others. Many more are approaching publication, including the relativity functions, the inverse tangent and such higher functions as the confluent hyper-geometric functions and Bessel functions for complex arguments and of fractional orders.

Tables are also being computed for the Bessel functions of the second kind with complex arguments, for Jacobi's elliptic functions with real and imaginary arguments and for the Tchebysheff polynomials, as well as for the first ten powers of the reciprocals of the integers from 1 to 1000 and other tables.

Even this list does not exhaust the ambitions of the American workers. Under active consideration are tables of the gamma function for complex arguments, $(x+iy)^n$ and others. A complete list of all tables published or as yet contemplated is given in a brochure which shows that this number is already no less than 30. Moreover, the Administrator (Irving V. A. Huie, 70 Columbus Avenue, New York City, N.Y.) invites suggestions for the computation of further tables.

The Technical Director of the Project, which is sponsored by the Bureau of Standards, is Arnold N. Lowan, and the technical staff consists of M. Abramovitz, G. Blanch, A. Hillman, W. Horenstein, I. Rhodes and H. E. Salzer. The tables are sold through the National Bureau of Standards, Washington, D.C.

There can be no doubt that the whole project will be of the utmost value to science, not only in the country of publication, but all over the civilized world.

*[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]*

XI. Rational Electrodynamics.—I. The Limitations of Classical Electromagnetism.

By E. A. MILNE, F.R.S.*

[Received October 20, 1942.]

1. IN two recent papers the author constructed a formulation of electromagnetism which, whilst coinciding in general consequences with a wide tract of the classical electromagnetic theory of Faraday and Clerk Maxwell, of Larmor and Lorentz, yet differed markedly in the ideas in which it originated. The papers cited were concerned mainly with the technicalities of the work, and with some of the consequences over cosmical periods of time and cosmical distances. In the present series of papers it is proposed to treat more closely the logical ideas lying behind the new theory, and to apply it more closely to particular problems. In particular, it will be shown how the *point-singularity* that is an electron behaves as though possessing a definite radius, the classical "radius of the electron," in spite of the fact that on the new theory the inverse square law of Coulomb holds down to indefinitely small distances. It will be shown further how an analysis of the possible orbits of two charges of opposite sign admits of the existence of a *neutron*, as well as the known Bohr orbits; this will lead to some considerations on intranuclear dynamics. The progress that will be made is intimately bound up with the construction of an electrodynamics in which the fundamental laws and equations are not mere inductions from macroscopic laws but should be valid at particle-velocities indefinitely close to that of light: the electrodynamics is linked at each stage with high-velocity rational dynamics. It is also bound up with a proper epistemology of the electromagnetic field, and the explicit mention, in every context, of both the observer observing the phenomena and the test-charge through which the field is known. Lastly, it will be shown that the new theory removes some of the difficulties that have gathered round the theory of Faraday and Clerk Maxwell.

A preliminary announcement of some of the results, without proof, was communicated to the Congress of Physicists held at Warsaw in the summer of 1938, and published in the record of that conference. But it was evidently very little understood by those present. It must be emphasized that the ideas about to be presented form a clean break with the traditional method of approach to a question of physical science—the method of inductive generalization from experiment. The new ideas are logical, not physical; but they are none the less philosophical in the

* Communicated by the Author.

literal meaning of that word, and they legitimately form a part of natural philosophy. They constitute the extension to the domain of electromagnetism of the methods which I have been developing since 1932 in the realms of kinematics, dynamics and gravitation. These methods have been much criticized by those who dislike them. But the most hostile critics, those who most resent the possibility of physical science becoming a purely deductive branch of knowledge, have not succeeded (a) in finding errors in the results of the methods, (b) in pointing to places where empirical appeals are actually made. Naturally, and legitimately, the course of the various investigations has been guided and influenced by our knowledge of empirical science, but this is a totally different matter from making, in the course of the deductions, specific empirical appeals, which are everywhere avoided.

2. One preliminary remark may be made to clear the air. A large tract of the ground in kinematics, dynamics, gravitation and electrodynamics can be dealt with by starting straightforwardly with the Larmor-Lorentz-Einstein transformation-formulæ of "special" relativity. But if we contented ourselves with so starting, there would be from the outset an empirical element, namely the empirical assumptions of a physical character that led Larmor and Lorentz on the one hand (the irrelevance of uniform relative velocity to the form of the equations of electromagnetism) and Einstein on the other hand (the constancy of the speed of light to all observers) to the isolation of these remarkable formulæ. In developing my own work, so many unexpected results followed from these formulæ, in deductive fields previously supposed to be the province of empirical physics, that it became desirable to examine whether there was indeed an empirical element in these formulæ of Lorentz. Close investigation showed that the only empirical appeal was to the *before-and-after* relation between events in the observers' consciousness; that the formulæ could be derived solely from the observers' acts of reading their own and one another's "clocks," a "clock" being merely a correlation between the real numbers and the events of the observer's consciousness; and that temporal observations alone appear to afford a basis for a large tract of deductive physical science.

Once the observer is armed with a clock (to the graduation of which, and its comparison with other observers' clocks, Dr. Whitrow and I have paid careful attention in our analysis) the use of a standard metre for measuring lengths becomes irrelevant. Lengths defined by reference to the standard metal are not lengths "operationally" defined, in Bridgeman's sense; they afford no basis for further deduction; and it is precisely because the appeal to the standard metre, made by Einstein in his proof of the Lorentz formulæ, is superfluous that he had later to introduce his empirical postulate of the constancy of the speed of light. In Einstein's theory, the postulate in question is the nexus between length- and time-measures, and is only necessary because length- and time-measures are there introduced independently. In my own work

(later in collaboration with Dr. Whitrow) it has been shown that the very observations which are necessary to fix the *epoch* of an event at a distance automatically suffice to fix a measure of the *distance* of that event. Not to use this measure of distance would be to reject the half of the information provided by the measures necessary to determine the epoch of a distant event, and so to place a barrier to the further progress of a purely deductive treatment. Critics have failed to recognize the essentially *logical* nature of this procedure for fixing measures of distance; they have failed to recognize the essential power of the newer methods because they have not freed themselves from the shackles of older, more empirical, methods. Critics have asked how these measures of length based on time-observations are ultimately identified with measures based on the standard metre. The answer to this has always seemed to me to be so obvious as scarcely to be worth mentioning: it is that, since we are in due course led to the same Lorentz formulæ as those to which Einstein was led by his empirical considerations involving the standard metre, the distance as measured on the new scheme must coincide with that based on the standard metre, *provided the measure of time used in the Lorentz formulæ is identical with that used in the present deductive treatment*. The whole question thus turns on the identification of the standard of time-measure, and this involves in turn the construction of a dynamics and its comparison with empirical dynamics. It subsequently appears that the standard metre is not of exactly "constant length" when the latter measure is based on the scale of time valid for the Lorentz-formulæ, namely what I have called "*t*-time," but is slowly expanding. However, the rate of expansion, *in that scale of time*, is so slow (the nebular recession-rate, 1 part in 2×10^9 per year) that it may be considered as only an ultimate refinement; to all practical intents, the measure of length defined by time-measures on clocks for which empirical and deductive dynamics coincide itself coincides with measures made with the standard metre.

Thus Kinematical Relativity is based on temporal observations alone. It is so based because, in the first place, it is found in this way to afford a rigorous and purely deductive proof of the Lorentz formulæ. In the course of this deduction, the theory is led to widely-flung generalizations of the Lorentz formulæ to observers in *non-uniform* relative motion, but these generalizations do not here concern us. To recapitulate: length-measures are introduced through the medium of time-determinations which an observer can in principle make by means of a clock at himself. The subsequent identification of the length-measures so found with those based on the standard metre is not an empirical appeal in the deductive theory *per se*; it is *not* used to enable us to make further deductions; it is merely part of the necessary dictionary of translation of the theory into terms that already have a meaning in physical science.

But there is a second and more fundamental reason why Kinematical Relativity is based on clock-readings made at instants determined by light-signals. That is, that an observer O cannot discuss a particle

somewhere in space until he knows it to be present. The simplest way of "knowing it to be present" is to look at it. In ordinary life we rarely "look at" things for an *instant*; we view them continuously illuminated for more or less considerable intervals. Now in our abstract scheme of relations which we are engaged in constructing, we want to refine this process down to viewing a particle *at an instant*, and the simplest way of doing this is to uncover a light at one's self, noting the time by one's own "clock" at which one does so, and then to note the time, by the same clock, at which the particle comes into view. This is, of course, the epoch of return of the "light-signal reflected from the particle," to use the language of sophisticated physics. But we are assuming no physics, consequently we are not bound to use its language; indeed, we may not do so, at this stage. We simply uncover the light, and note when we first see the particle. No more elementary act of perception, no more immediate sense-datum, or, rather, relation between two sense-data, can be conceived. Thus our abstract scheme pays due and full regard to the theory of knowledge, as this has been developed by philosophers from Locke to Russell. We *assume* no properties of light whatever; we make no appeal to the physics of light; we do not need even to assume that "light" exists. We merely analyse the process of taking an instantaneous view of a particle—exactly as if we were photographing it by flashlight. The only particles we discuss in these papers are particles whose presence is perceived by this mechanism, and of which their observers thus became aware. There is no question of introducing any assumption about "uncertainties" in "conjugated dynamical variables." We cannot yet speak of a pair of such variables, for we have not yet constructed them; we have time-measures only.

The epochs t_1 and t_2 , of uncovering the light and seeing the particle, read on the observer's own clock, afford t , the conventional epoch at which the particle was illuminated, and x , its equally conventional distance at this epoch, by the relations

$$t = \frac{1}{2}(t_2 + t_1), \quad x = \frac{1}{2}c(t_2 - t_1),$$

where c is arbitrary. For convenience, once one observer has selected a value for c , all other observers select the same value. A consequence of this is that, however the clock of an observer runs, he always attributes the speed c to the assumed "signal"—if he cares to assume propagation of a signal. Actually all assumption of *propagation* of a signal is irrelevant to the theory, as explained previously. We introduce nothing but the elementary act of seeing, combined with the observer's awareness of the passage of time at himself.

I must make it clear that this "act of seeing" is on a different level from the appeal to experience in the observer's awareness of the passage of time at himself. The act of seeing is the condition of having anything external to one's self to discuss at all. Without it, we should be purely in the position of a solipsist Descartes. But the awareness of the before-and-after relation, the possibility of arranging experienced events in a

purely one-dimensional or "time"-order, is a thing which might have been otherwise, something which a philosopher calls "contingent"—a fact of empirical experience, therefore, in the true philosophic sense.

One trivial remark before we pass on: it is convenient to retain the symbol c throughout, explicitly, in all our formulæ. This is contrary to the current practice of relativists, which is to put $c=1$. To do this, however, would be to obscure the relationship and relative importances of different terms; and it would lead to the non-recognition of the genuine absence of c from certain formulæ and equations. To retain c gives a certain compact "stratification" to formulæ; and it avoids the considerable trouble of restoring it when it is needed.

3. It is convenient to begin this series of papers proper with an account of the limitations of classical electromagnetism. These are not reviewed in any spirit of hostility to classical electromagnetism. But it is important to avoid the attitude that classical electromagnetism (or, too, "general" relativity) are sacrosanct, above criticism—that it is an impiety to lay hands on them. Actually many of the difficulties encountered by modern physical theory are due to the forgetting that classical electromagnetism was an empirical theory, founded on generalizations from large-scale phenomena, and utterly unsuitable for a deep-lying analysis of microscopic phenomena. Such an analysis cannot be undertaken by any patching-up of the classical theory, such as "quantizing" the classical theory; it requires a reconstruction of the basal elements of the subject *de novo*. To such we now proceed.

I. The Limitations of Classical Electromagnetism.

4. The classical theory of electromagnetism originated in the physical ideas of Faraday, given mathematical expression by the genius of Clerk Maxwell. They were developed into a *theory of electrons* by Larmor, Lorentz and others. Notable contributions were made by Poynting and by Wiechert. It is essentially a *field theory*, in which the "state of the field" is supposed to have a physical reality. In this theory the electron is contemplated as a very small element of charge, but never as a point-singularity; even when it is "reduced to a point" it is merely as the limit of a volume-element. The theory is given mathematical embodiment by means of "field equations" in empty space; these are generalized in the presence of matter, but when matter is considered as having a purely electrical constitution the charge is represented by the addition of terms denoting a volume-density. The field in empty space is characterized by two vectors \mathbf{E} and \mathbf{H} , the electric and magnetic intensities, whose interplay, expressed in Maxwell's four equations, gives rise to the phenomena of the field. The mechanical effects of the field on massive particles are attributed to the Larmor-Lorentz ponderomotive force $\mathbf{E} + (\mathbf{V} \wedge \mathbf{H})/c$ where \mathbf{V} is the velocity of the moving particle, and c is the speed of light. The intensities \mathbf{E} and \mathbf{H} are defined through the forces experienced by elementary electric charges and magnetic poles respectively. In its inception the theory is built up out

of electrostatic theory, with the inverse square law of Coulomb as its foundation, and a similar magnetostatic theory. These two branches were then fused together by Faraday and Maxwell, so that the time-rate of change of \mathbf{E} or \mathbf{H} in a non-static field determined \mathbf{H} or \mathbf{E} respectively.

5. Superb as this structure is as a means of analysing many physical phenomena in nature, it is far from satisfying logically. The first difficulty is concerned with the location of *energy* in the field. Energy is an abstract concept derived from the calculation of the performance of mechanical work. The energy associated with a pair of charges is of course the work, negative or positive, that can be extracted from them by suitably displacing them. This is a perfectly definite quantity: namely the energy of two point-charges e_1 and e_2 at distance r is $e_1 e_2 / r$, assuming the inverse square law of Coulomb. A single elementary point-charge by itself has no energy, because no work can be extracted from it, since by calling it elementary we mean that it cannot be further split up. Now when the charges in the field, instead of being discrete elementary point-charges $e_1, e_2, \dots e_n$, can be represented as continuous surface-distributions of finite surface-density σ , or continuous volume-distributions of finite volume-density ρ , the total energy represented by an expression of the type $\sum_{r,s} e_r e_s / r$ can be expressed via a potential function, and the use of Green's theorem, as the volume-integral of $\mathbf{E}^2 / 8\pi$ throughout all space occupied by "tubes of force," and since there is no field in the interiors of conductors, this comes to the same thing as saying that the energy of the system of surface- and volume-charges can be regarded as distributed throughout all space at the rate of $\mathbf{E}^2 / 8\pi$ per unit volume.

Now the transformations used in applying Green's theorem are impossible when part of the charge consists of actual point-charges. Thus if the theory is to give an account of electrons as possible point-singularities, it must eschew this representation of energy as a volume-integral. Forgetfulness of this simple matter has often led to the statement that "the self-energy of a point-charge is infinite," and various *ad hoc* modifications of the fundamentals of the theory have been suggested by Born and others in an endeavour to remove this supposed impasse, and to make the self-energy of a point-charge finite. Actually an elementary point-charge has no self-energy. If it can be split into parts and separated it must have a structure, and cannot be an indivisible elementary point-charge such as an electron is supposed to be; if it is not capable of being decomposed into parts, then it cannot have any energy ascribed to it. The mere fact that $\mathbf{E}^2 / 8\pi$ integrated throughout the space surrounding a point-charge is infinite is not of physical significance as regards energy.

6. The difficulties into which we are led by ascribing energy to volume-elements of the field are of course of no importance so long as the charge distribution is not viewed microscopically. For example, localization of energy in the field will be valid when conductors may be supposed to be

the seat of surface-distributions of charge, or in the case of the space-charges occurring, for example, in a valve. But the difficulties become dominant when we envisage the micro-structure of the charge-distribution, when this contains charges as point-singularities.

Once we recognize the illegitimacy of the localization of energy in the field in the presence of point-charges, we can no longer deduce Poynting's theorem. Consequently calculations of the radiation emitted by accelerated point-charges by the usual formula became invalid. As an example, the calculation of the field of a moving electron by Wiechert's method becomes impossible. For in that method the varying contributions to the field of a small but finite-density volume-charge are determined, the volume-charge then regarded as indefinitely small and the radiation calculated by Poynting's formula. If the electron were in fact a small volume-distribution, it would need to be held together by other than purely electromagnetic forces, and we could not then discuss the field and the motion of the charges in it on purely electromagnetic grounds. This has often been pointed out. But what is important here is that if on the contrary an elementary charge is a pure singularity, it is something essentially different from the limit of a small volume-element, and the calculation of the radiation emitted by it, by Wiechert's formula and Poynting's theorem, fails entirely. Hence we can have no confidence in Larmor's formula for the rate of radiation, $\frac{2}{3}e^2f^2/c^3$ (f being the acceleration) as applied to a point-charge.

7. This formula does not in any case cover the facts of experience as applied to micro-phenomena. It fails, as is well known, to give non-radiating orbits in atomic stationary states. Further (what is less frequently noticed) it fails to account in a classical way for the well-established fact of the *reversibility* of the emission of radiation in the act of absorption. For, giving a rate of radiation proportional to the *square* of the acceleration, it is not reversed in sign on reversing the acceleration. The observed reversibility of the emission and absorption of radiation on transitions between stationary states strongly suggests that the rate of radiation should be proportional to the first power, or at least an odd power, of the acceleration.

8. We turn to magnetic phenomena. In the classical theory of electromagnetism the magnetic intensity is never properly introduced. It is taken over from magnetostatic fields, in which the magnetic intensity is defined through the intervention of a hypothetical isolated magnetic pole. But isolated magnetic poles do not exist when we take a *micro-view* of the field, and magnetic intensity consequently needs redefining. The natural way to introduce magnetic intensity is as a constituent of the mechanical force on a moving small charge. But we must be careful here not to proceed in a naïve way, for the relative motions of the test-charge and the other charges in the field will by themselves introduce additional magnetic fields, naturally depending on the velocity of the test-charge. These will in turn introduce additional constituents of the mechanical

force on the moving test-charge, and great circumspection will be required to isolate an objective magnetic field, if indeed it is possible at all, by this method. It is of course precisely in the field of micro-phenomena that the application of classical theory has encountered grave difficulties. The avoidance of these has led to the introduction of such non-classical concepts as that of electron-spin, of which in my opinion no rational account has yet been given, and to the arbitrary introduction in sundry places of a factor of $\frac{1}{2}$.

9. The classical theory can scarcely be expected to be valid, save as it were accidentally, for high-speed charges, *i. e.* for charges moving with speeds comparable with that of light. For electric intensity is defined electrostatically, that is, by the mechanical effects on macro-charges at rest. We do not in fact know that the "force" on a moving charge, in a given electric field, is independent of the velocity of the charge. In any case, "force," in the context of high-speed motion, requires very careful definition, and must be considered in connexion with relativistic equations of motion; and Einstein's "special relativity" dynamics, in which energy is the fourth component of a 4-vector of which the other components are momentum, is by no means the only type of relativity dynamics that can be constructed. It is small wonder, then, that the electrodynamics of the nucleus of the atom has made such little progress, for nuclear dynamics is essentially the dynamics of particles which, in consequence of their great mutual proximity, must be moving with speeds which are a large fraction of that of light.

10. The classical theory of the electron, on certain not very convincing assumptions, ascribed to it a radius of the order of e^2/mc^2 . The order of magnitude of this expression undoubtedly gives the order of the distance from an electron at which the Coulomb law *appears* to break down; and it is of the order of magnitude of the nucleus. Although the supposed proof of the radius of the electron being e^2/mc^2 is no longer accepted, there has recently been a movement to introduce this length as giving a fundamental aspect of the structure of the electron, without further particularization. This procedure is somewhat mystical, and does nothing more than conceal our lack of understanding of the mode of behaviour of elementary charges. A rational theory must be expected to exhibit this length-quantity e^2/mc^2 emerging of itself, without specific assumptions. But in this respect classical theory is again in a dilemma: for, once again, either the electron has a structure, in which case other than purely electromagnetic forces are required; or the electron is a point-singularity, in which case there would appear small prospect of a fundamental length e^2/mc^2 emerging, for there would be apparently no reason for the "breakdown" of the Coulomb law at any particular distance.

11. Lastly, the classical theory of electromagnetism is supposed to apply to "empty" space, without further detailed specification of the

observer or his frame of reference. The frame of reference is intended to be an unaccelerated or "inertial" frame. But no test is given or mentioned by which an observer may recognize which of the possible frames of reference available to him are inertial frames. It is of little use to say that he must choose an unaccelerated frame. Unaccelerated relative to what? If we have one class of frames, $F_1 \dots F_n$, moving with uniform relative velocities V_{rs} , and a second class of frames, $f_1 \dots f_n$, moving with uniform relative velocities v_{rs} amongst themselves, but all moving with acceleration a relative to the first set, how is the observer to choose which set constitute an inertial frame? This difficulty is not peculiar to electromagnetic theory—it occurs in general dynamics; Einstein and Infeld stated in 'The Evolution of Physics' that no one had yet succeeded in defining an inertial frame. But in a subject of which the *motion* of charges is the essence, one would expect no satisfying theory until this difficulty is met.

12. None of the foregoing criticisms must be taken to detract from the admiration which every physicist feels for Faraday's and Maxwell's superb creation. Nevertheless, the resulting theory—an induction, as stated above, from large-scale phenomena—is by the nature of the case quite unfitted as a tool for investigating microscopic phenomena. The elucidation of these requires a new complex of ideas. Such a complex is afforded by kinematic relativity. The difficulties encountered by the classical theory, very briefly summarized above, cannot be removed by going over them one by one and trying to modify the theory so as to surmount each in turn. Rather must we make a new starting-point, and see how the new theory fares. It will be found that the various difficulties—energy localization, reversibility of emission of radiation, magnetic intensities, "radius" of the electron, high-speed dynamics and selection of inertial frames—are all naturally met by the methods of kinematic relativity without special assumptions. But of course the ideas of kinematic relativity, hitherto confined to kinematics, dynamics and gravitation, need to be deepened before they can be made useful in the field of electrodynamics.

Bibliography.

The ideas of Kinematic Relativity and the basing of measures on temporal observations were developed by the author in a series of papers and lectures which up to 1935 were summarized in :

E. A. Milne, 'Relativity, Gravitation and World-Structure' (Oxford, 1935).

The full theory of the "equivalence" of observers and the nature of time-keeping, clock correlation, etc., was further developed in :

E. A. Milne and G. J. Whitrow, "On the Meaning of Uniform Time, and the Kinematic Equivalence of the Extra-galactic Nebulae," *Zeits. für Astrophys.* xv. pp. 263–298 (1938).

After the publication of 'World-Structure,' the ideas were applied more particularly to the abstract theory of dynamics and gravitation in the following series of papers by the present author :

- "On the Foundations of Dynamics," Proc. Roy. Soc. A, cliv. pp. 22-52 (1936).
- "The Acceleration Formula for a Substratum," Quart. Journ. Math. (Oxford), viii. pp. 22-31 (1937).
- "Kinematics, Dynamics and the Scale of Time," I., II. and III., Proc. Roy. Soc. A, clviii. pp. 324-348 (1937); clix. pp. 171-191 and 526-547 (1937).
- "The Inverse Square Law of Gravitation," I., II. and III., Proc. Roy. Soc. A, clvi. pp. 62-85 (1936); clx. pp. 1-23 and 24-36 (1937).

The ideas on electromagnetism which are more particularly the subject of the present series of papers were developed in :

- "On the Equations of Electromagnetism.—I. Identifications; II. Field Theory." Proc. Roy. Soc. A, clxv. pp. 313-332 and 333-357 (1938).

General, more popular accounts of the work (excluding electromagnetism) were given in :

- "Kinematical Relativity" (Presidential Address to the London Mathematical Society), Journ. London Math. Soc. xv. pp. 44-80 (1940). (This contains a fuller bibliography to that date.)
- "Cosmological Theories," Astrophys. Journ. xci. pp. 129-158 (1940).
- "Remarks on the Philosophical Status of Physics," Philosophy, xvi. pp. 1-16 (October 1941).

XII. *Rational Electrodynamics.*—II. *The Ideas of Kinematical Relativity.*

By E. A. MILNE, F.R.S.*

[Received October 20, 1942.]

13. The central ideas of kinematical relativity cluster round the following situation. (I content myself with stating the situation without proofs—proofs have been given in detail in the papers cited in the Bibliography.)

Consider an expanding swarm of particles, of which a typical member has a position-vector \mathbf{P} and velocity \mathbf{V} at epoch t , as measured by an observer O at the origin. The epoch t is reckoned from the natural zero of time, and the assignments of position and velocity have been made by O by means of light-signals recorded on his clock. Let each particle move with uniform velocity \mathbf{V} , given by the relation

$$\mathbf{V} = \mathbf{P}/t, \quad (1)$$

* Communicated by the Author.

and let the velocities of these particles be distributed according to the velocity-distribution law

$$\frac{B du dv dw}{c^3(1-\mathbf{V}^2/c^2)^2} \dots \dots \dots (2)$$

Then the positions of the particles will be distributed according to the particle-density law

$$\frac{Bt dx dy dz}{c^3(t^2-\mathbf{P}^2/c^2)^2} \dots \dots \dots (3)$$

Here, u, v, w are the components of \mathbf{V} ; x, y, z the components of \mathbf{P} , in any non-rotating frame attached to O .

This distribution of particles is called a *substratum*. It has the property that if the description of the swarm is transferred from O to any other particle O' , a member of the swarm, by means of the Lorentz transformation, so that what was described by O as \mathbf{P} or (x, y, z) , \mathbf{V} or (u, v, w) , t , becomes (to O') \mathbf{P}' or (x', y', z') , \mathbf{V}' or (u', v', w') , t' , then the new description in terms of the primed co-ordinates (namely, the co-ordinates used by O') is precisely of the form (1), (2), (3) as before. In this description c is a conventional constant, chosen by agreement between all observers of the type O or O' . and B is another conventional constant. The constant c proves to represent the speed of light in the measures used by any observer; the meaning of B we expound later. The legitimacy of employing the Lorentz formulæ is established by a specific investigation which, starting from the definitions of the epoch and distance measures used by O and O' , together with the observations O and O' can make on one another, succeeds in deducing the Lorentz transformation without any physical assumptions.

Any observer O' attached to a particle of the system is called a *fundamental observer*. Since O is central in his own description of the system, so is any O' in *his* description. Every particle is therefore central to the particular observer associated with it. To each such observer the swarm appears to be confined to the interior of an expanding sphere, of radius ct , towards the boundary of which the particle-density becomes infinite. The finite expanding sphere has, however, to O all the properties of infinite space, for no emissary he can send out can ever reach its boundary; for no emissary can reach the velocity c , and so any emissary perpetually fails to overtake the boundary. Moreover, space "outside" the expanding surface $|\mathbf{P}|=ct$ has no meaning for any O , since he cannot see into it, owing to the infinite spatial density at $|\mathbf{P}|=ct$. Actually we do not suppose infinite spatial density, for the swarm is confined to the *interior* of $|\mathbf{P}|=ct$, and so constitutes an *open set* of points.

The various observers O are said to be equivalent in pairs since they make reciprocally identical observations on one another; we write this equivalence $O \equiv O'$. The observers O and O' are equivalent in another sense, in that they make identical observations on the whole system of which they are members; we write this equivalence $O \equiv O'$. It may be

mentioned that each O uses the ordinary flat Euclidean space of physics ; but these observers' spaces are *private* spaces, for the world-wide space of O at one instant t in his world-wide assignment of epochs does not transform into the world-wide space of O' at some instant t' in *his* assignments. The current conventional method of describing the spatio-temporal relations of O and O' is to describe them as traversing diverging world-lines in a common space-time. But space-time as an aid to the scientific imagination is more bother than it is worth ; I am writing for physicists, who use as a rule ordinary flat space, and I shall find it convenient to avoid all mysticism by foregoing any mention of the construct "space-time." (I occasionally use the phrase for *mathematical* convenience only.)

The swarm of particles reduced to a *point*, $\mathbf{P}=0$, at the epoch $t=0$, which constitutes a natural origin of time for the swarm. Negative epochs t have no meaning. It is always understood that any stated epoch t is always reckoned from this natural zero of time. If the system is considered as *beginning* at $t=0$, then t represents the "age" of the swarm, to O, at the event considered.

14. So far the system has merely been defined *kinematically*. It is a set of constructed motions. The fascinating dynamical question then presents itself : if once constructed, will it go on in this prescribed kinematic motion of itself, of its own accord ? That it is likely to do so is seen from the fact that each particle is central, in its own estimation, and hence has no tendency to do other than remain central.

But we can tackle the question on a much deeper basis. Suppose a *free* particle is launched from any member of the system, at any assigned epoch, at any assigned velocity, as reckoned by O. What will be its motion ? Is it possible, without any appeal to any physically observed or empirically known law of motion, to infer its motion, that is, its instantaneous acceleration, simply from a logical analysis of its relation, as a free particle, to the swarm of given particles amongst which it is moving ?

15. The answer is that it is so possible. This is a completely revolutionary state of affairs, making a sharp break with the previous tradition in occidental physics. Without any experimental appeal, we can infer that the motion of a *free* particle at the position \mathbf{P} , moving with velocity \mathbf{V} at the instant t (all as reckoned by O) is given by

$$\frac{d\mathbf{V}}{dt} = -\frac{\mathbf{Y}}{\mathbf{X}}(\mathbf{P}-\mathbf{V}t), \quad \dots \dots \dots (4)$$

where $\mathbf{X}=t^2-\mathbf{P}^2/c^2, \quad \mathbf{Y}=1-\mathbf{V}^2/c^2. \quad \dots \dots \dots (5)$

Moreover, the equation of motion (4) becomes of the same form when transformed into the co-ordinates used by any other observer O'.

16. It will be useful at this stage to mention a point of technique. Though we do not use modes of expression involving "space-time," we

need a convenient way of readily verifying that expressions, equations, &c., remain covariant or invariant in form under a Lorentz transformation from O to O' . To do this we note that

$$X \quad \text{and} \quad Z/Y^\dagger,$$

where X and Y are as in (5) and Z is given by

$$Z = t - \mathbf{P} \cdot \mathbf{V}/c^2, \quad \dots \dots \dots (5')$$

are invariants, *i. e.*, 4-scalars, under a transformation from any O to any O' ; whilst

$$(\mathbf{P}, \quad ct)$$

and

$$\left(\frac{\mathbf{V}}{Y^\dagger}, \quad \frac{c}{Y^\dagger} \right)$$

are 4-vectors under the same transformations. We can now rapidly see that (4) is unaltered in form on transforming from O to any O' . For scalar multiplication by \mathbf{V} gives at once

$$\frac{\mathbf{V}}{c^2} \cdot \frac{d\mathbf{V}}{dt} = -\frac{Y}{X} (tY - Z), \quad \dots \dots \dots (4')$$

whence, differentiating out $Y^{-\frac{1}{2}} \frac{d}{dt} \left(\frac{\mathbf{V}}{Y^\dagger} \right)$ and using (4) and (4'), we have

$$\frac{1}{Y^\dagger} \frac{d}{dt} \left(\frac{\mathbf{V}}{Y^\dagger} \right) = -\frac{1}{X} \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right). \quad \dots \dots \dots (6)$$

Since $Y^\dagger dt$ is a 4-scalar, and \mathbf{V}/Y^\dagger and \mathbf{P} are the space-parts of 4-vectors, and since lastly Z/Y^\dagger is a 4-scalar, the above equation is the space-part of a 4-vector equation of which the time-part is

$$\frac{1}{Y^\dagger} \frac{d}{dt} \left(\frac{c}{Y^\dagger} \right) = -\frac{1}{X} \left(ct - c \frac{Z}{Y} \right), \quad \dots \dots \dots (6')$$

which follows from (4'). The steps by which (6) was derived from (4) can now be reversed, and this shows that (4) conserves its form under the transformation from O to any O' .

17. We return to (4). It shows that if, at any instant, $\mathbf{V} = \mathbf{P}/t$, then at that instant $d\mathbf{V}/dt = 0$, and so \mathbf{V} continues constant, *i. e.*, $\mathbf{V} = \mathbf{P}/t$ continues to be obeyed. It follows that the swarm of particles defined by (1), (2), (3) continues of itself, and constitutes not only a *kinematically prescribed* system but a *dynamical* system. By sufficiently deep discussion of the nature of the motion of a *free* particle in the presence of a swarm of particles of *prescribed* motion, we have been able to show that the swarm itself would behave in the same way *naturally*, if once started. That is to say, we have established a *dynamical* result from discussion of a *kinematic* situation, by a purely logical process. To cross the frontier between kinematics and dynamics is as decisive a step as to cross the frontier between the non-living and the living. The possibility of doing this, thus realized, is in sharp contrast to the natural philosophy of Renaissance physics.

There is clearly something very significant about a situation in which this step is possible. It would have been a scientific crime on my part to have failed to develop further a situation so promising of further results, gleaming with such a strange new light, and accordingly since 1932 I have put it in the forefront of my researches. The fruitfulness of the resulting investigations, and the many surprising turns they have taken, have fully justified this policy.

Once we have established that the substratum or simple kinematic system (1), (2) and (3) is a *natural* system, it suggests itself that we enquire as to the course of other phenomena which may take place in its presence, and ask whether we can elucidate and describe them, again without making appeals to empirically-known facts. Such phenomena will include those of gravitation and electromagnetism.

18. We first notice that the system (1), (2), (3) supplies just the set of equivalent frames of reference that "general" relativity was unable to select. "General" relativity demands that laws of behaviour of natural systems shall be describable in the same way in *all* sets of co-ordinates, *i. e.* in the co-ordinates used by *all* observers in *all* frames of reference. This is too stringent a demand. There is no reason why the descriptions of natural phenomena, by observers using different types of frames of reference, using co-ordinates constructed out of observations combined in different kinds of ways, should be capable of being thrown into the same form. What *is* necessary is that observers who stand in the same relation to the system under observation and who make their observations of the system and construct therefrom their co-ordinates by the same rules, should be led to descriptions of the phenomena in question of the same form. Now in (1), (2) and (3) each observer O is the centre of the expanding swarm, to himself, and his description of its unfolding history, in space and time, starting from his natural zero of time, is identical with that of every other observer O; he stands in the same relation to the system as any other O. Moreover, the rules and conventions by which he attributes co-ordinates to distant events are the same as for all the other equivalent observers. Consequently the description of the rules governing the course of any phenomenon should be the same for every O.

An observer O is one moving with a fundamental particle of the swarm. We can consider other observers, moving in any other way whatever, and examine their descriptions of phenomena. They will be equally valid descriptions, but will not coincide in form with those of the fundamental observer O. There will indeed be classes of other observers, say Q, whose descriptions will be of the same form amongst themselves. We impose no conditions on such observers' descriptions. The only observers about whose descriptions we can predicate that they are identical in form with O's description are this observers' "equivalent" to O, *i. e.* those moving with the equivalent particles of the swarm.

19. It will now be apparent that kinematic relativity is more general

than "general" relativity in that it imposes less stringent restrictions on possibility. Instead of requiring that *all* systems of co-ordinates yield the same description, it requires this only of a strictly limited set of observers, whose relation to the complete system imposes this requirement of necessity.

20. This special set of observers, with their systems of co-ordinates, *i. e.* frames of reference, have been called by some critics "privileged" observers. But they enjoy no special privileges they are not entitled to. We cannot deny that from the identity of their relationships to the whole system of which they are part they must be led to similar descriptions of relations between events, *i. e.* to similar laws of phenomena. We are imposing weaker requirements than general relativity, and it is *a priori* possible that these weaker requirements are not sufficiently strong to determine the descriptions of phenomena. Actually these weaker requirements are just sufficient, as we shall see.

21. This set of frames of reference, *i. e.* of observers with sets of co-ordinates determined from similar observations by similar rules of combination, is now appropriate for the investigation of electromagnetism. The observers are now associated with defined frames of reference which are relatively unaccelerated. Each observer counts himself as at rest, and the other observers in motion. Each observer can identify his own frame of reference with an inertial frame.

This set of frames of reference, constituting what we have called a substratum, was originally developed to represent the phenomenon of the expanding universe of receding extra-galactic nebulae, each nebular nucleus being the representative of some fundamental particle. But the system of separating observers has a far wider significance. Just as in plane geometry we need a plane 2-spread as the seat of the structures constructed, so in dynamics, gravitation and electromagnetism we require a 3-dimensional continuum containing temporal relations between its observer-members (particle-members replacing the "points" in a geometry) as the seat of the phenomena to be considered. We can make no progress by considering phenomena in a vacuum, in empty space. By beginning with our system of observers and their frames of reference, all explicitly mentioned, we construct the necessary scaffolding for giving access to phenomena. We cannot have scenery without a scene—without a stage. The substratum is a suitable scene for the phenomena of gravitation and electrodynamics.

22. Having obtained the equation of motion of a free particle, in the form of (4) or (6), we are in a position to construct a *dynamics*; this is a necessary step in the construction of an *electrodynamics*. The construction of a dynamics can be undertaken in more than one way, but in order that the constructed dynamics shall be useful it is necessary that the terms used represent entities subsequently identifiable with their counterparts in empirical dynamics.

A dynamics based on equation (4) or (6) in the presence of the substratum has been constructed in detail in a series of papers in the 'Proceedings of the Royal Society.' Here it suffices to state the leading results without proof.

23. The equation of motion (4) or (6) is utterly unlike the equation of motion of a free particle in Newtonian dynamics. It must be remembered that it is relevant to the scale of time in which t is measured. As it stands, the term on the right-hand side can be interpreted as the *gravitational pull* of the substratum, in this scale of time, on the particle at P , moving with velocity V at epoch t . Its "physical" setting is seen from the fact that the acceleration is directed towards the point Vt , which, being the fundamental particle at rest relative to the given particle, is the centre of the substratum to an observer moving with the free particle in question, and so the centre towards which the acceleration is naturally directed. We can effect an identification of one aspect of this acceleration with empirical gravitational theory which is useful in the sequel. For a particle at rest relative to O , at a small distance $|P|=r$ from O , the acceleration reduces to $-r/t^2$. If we assign a mass m_0 to each fundamental particle, since the particle-density at O , by (3), is B/c^3t^3 , the mass in the sphere of radius r is approximately $\frac{4}{3}\pi r^3 m_0 B/c^3t^3$, and its gravitational pull $-\gamma(\frac{4}{3}\pi r^3 m_0 B)/c^3t^3 r^2$, where γ is the usual "constant" of gravitation; this reduces to $-\frac{4}{3}\pi m_0 B \gamma r/c^3t^3$. Identifying the two determinations of the acceleration, we have

$$\gamma = c^3 t / M_0,$$

where

$$M_0 = \frac{4}{3}\pi m_0 B.$$

The interpretation of M_0 is that it is the mass which would be assigned to the matter occupying the sphere containing the substratum if the latter were filled with matter *homogeneously* at a density equal to the actual substratum-density at the observer; for this mass would be

$$\frac{4}{3}\pi (ct)^3 \times m_0 B / c^3 t^3 = \frac{4}{3}\pi m_0 B.$$

This may be called "the mass of the equivalent homogeneous universe." (It must be remembered, of course, that in the t -scale the actual substratum is not homogeneous.) The constants m_0 and B are purely conventional numbers, and only their product has a meaning. When we evaluate $c^3 t / M_0$ by means of astronomical data, using for t the present age of the universe determined by the observed recession of the nebulae, we obtain a value agreeing with the observed value of γ . (See § 43 below.)

The idea of a "constant" of gravitation varying with t , the age of the universe, seems at first sight strange and revolutionary, but it must be remembered that this description of γ is appropriate only to the scale of time t . We shall see shortly that this scale of time is not the scale of time which is appropriate to the Newtonian description of dynamics and gravitation. But the t -scale is the fundamental and natural scale for a basic analysis of gravitation.

24. If a particle is not moving according to equation (4) or (6) it is said to be acted on by a *force*. We can define this force by means of a measure of the extent to which (4) or (6) is not fulfilled.

It is readily verified that the invariant ξ defined by

$$\xi = Z^2/XY \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

is an integral of the equation of motion (4) or (6). It remains constant during *free* motion. We can thus write the equation of *free* motion in the form

$$\frac{1}{Y^\dagger} \frac{d}{dt} \left[m_{\xi^\dagger} \frac{\mathbf{V}}{Y^\dagger} \right] = -\frac{m_{\xi^\dagger}}{X} \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right), \quad \dots \quad (8)$$

where m is some positive constant. The invariant $m\xi^\dagger$ now plays the part of *mass*. We put $m\xi^\dagger = M$. For a fundamental particle, for which $\mathbf{V} = \mathbf{P}/t$, ξ reduces to unity, and M reduces to m_0 ; thus the mass is the same for all fundamental particles. It should be carefully noted that in this dynamics *mass* is an invariant, taking the same value, for any particle, whatever fundamental observer reckons it. For the fundamental observer coincident in position with the particle in question, $m\xi^\dagger$ reduces to $m/(1 - \mathbf{V}^2/c^2)^{\frac{1}{2}}$, since $\mathbf{P} = 0$. The fact that all fundamental particles have the same mass, in spite of their differing velocities, is the consequence of the two facts (a) that they are all equivalent, (b) that mass is an invariant.

Suppose now that the particle is *not free*. Then (8) is not obeyed—the two sides will be unequal. Write

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(m \xi^{\frac{1}{2}} \frac{\mathbf{V}}{Y^{\frac{1}{2}}} \right) = -\frac{m \xi^{\frac{1}{2}}}{X} \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right) + \mathbf{F}, \quad (9)$$

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(m_{\xi^{\frac{1}{2}}} \frac{c}{Y^{\frac{1}{2}}} \right) = - \frac{m_{\xi^{\frac{1}{2}}}}{X} \left(ct - c \frac{Z}{Y} \right) + F_t, \quad (9')$$

where $(\mathbf{F}, \mathbf{F}_t)$ is by its definition a 4-vector. Then we call $(\mathbf{F}, \mathbf{F}_t)$ the *external force* acting on the particle.

The work done by the external force will be defined as at a rate equal to the 4-scalar product of the external force and the velocity of the particle relative to its environment in the substratum. The latter velocity, for reasons we will not here go into, is represented by the 4-vector

$$\frac{\mathbf{v}}{\mathbf{Y}^{\frac{1}{2}}} - \mathbf{P} \frac{\mathbf{Y}^{\frac{1}{2}}}{\mathbf{Z}}, \quad \frac{c}{\mathbf{Y}^{\frac{1}{2}}} - ct \frac{\mathbf{Y}^{\frac{1}{2}}}{\mathbf{Z}}.$$

(That this is a plausible representation of this relative velocity is seen from the circumstance that for $|\mathbf{V}| \ll c$, $|\mathbf{P}| \ll ct$, *i. e.*, in ordinary experience, this 4-vector reduces to $(\mathbf{V} - \mathbf{P}/t, 0)$, as we should expect.) The following identity, derived from (9) and (9'), can now be verified :

$$\mathbf{F} \cdot \left(\frac{\mathbf{V}}{Y^{\frac{1}{2}}} - \mathbf{P} \frac{\mathbf{Y}^{\frac{1}{2}}}{Z} \right) - F_t \left(\frac{c}{Y^{\frac{1}{2}}} - ct \frac{Y^{\frac{1}{2}}}{Z} \right) = \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (mc^2 \xi^{\frac{1}{2}}). \quad (10)$$

The right-hand side is therefore equal to the rate of performance of work by the external force. Accordingly $mc^2\xi^{\dagger}$ may be taken as a measure of the *kinetic energy* E of the moving particle. We put

$$E = mc^2\xi^{\dagger}. \quad (11)$$

Thus in this dynamics, as in Einstein's, we have the relation between energy and mass

$$E = Mc^2. \quad (12)$$

But it is particularly to be noted that in our dynamics, unlike Einstein's, energy is an invariant, taking the same value for all fundamental observers, and not merely the fourth component of a 4-vector of which the space-part is the momentum-vector.

25. In the present dynamics, momentum \mathbf{p} is clearly represented by the 4-vector

$$m\xi^{\dagger} \frac{\mathbf{V}}{Y^{\dagger}}, \quad m\xi^{\dagger} \frac{c}{Y^{\dagger}},$$

but we have a 4-vector more nearly representing our ordinary notion of momentum if we consider the 4-vector whose space-component is the Hamiltonian expression

$$Y^{\dagger} \frac{\partial E}{\partial \mathbf{V}},$$

which comes to be

$$M \left[\frac{\mathbf{V}}{Y^{\dagger}} - \mathbf{P} \frac{Y^{\dagger}}{Z} \right],$$

whose time component is

$$M \left[\frac{c}{Y^{\dagger}} - ct \frac{Y^{\dagger}}{Z} \right].$$

This is just M times the 4-vector we have taken to be the velocity of the particle relative to its environment.

26. For future reference we record another identity analogous to (10), namely

$$\mathbf{F}_t \frac{c}{Y^{\dagger}} - \mathbf{F} \cdot \frac{\mathbf{V}}{Y^{\dagger}} = \frac{1}{Y^{\dagger}} \frac{d}{dt} (mc^2\xi^{\dagger}). \quad (10')$$

27. We shall naturally enquire next whether a force (\mathbf{F} , \mathbf{F}_t) introduced in this way is capable of being represented as the gradient of a potential χ , a function of \mathbf{P} and t . If we seek to set $\mathbf{F} = -\partial\chi/\partial\mathbf{P}$, $\mathbf{F}_t = +\partial\chi/c\partial t$, we rapidly find that (10) and (10') become inconsistent. Further consideration shows that we have to separate out from our term defining *force* a term due to the effect of change of mass with velocity, and that we have to write

$$\mathbf{F} = -\frac{\partial\chi}{\partial\mathbf{P}} + 2 \frac{\mathbf{V}}{Y^{\dagger}} \frac{1}{Y^{\dagger}} \frac{d}{dt} (m\xi^{\dagger}), \quad (11)$$

$$\mathbf{F}_t = +\frac{1}{c} \frac{\partial\chi}{\partial t} + 2 \frac{c}{Y^{\dagger}} \frac{1}{Y^{\dagger}} \frac{d}{dt} (m\xi^{\dagger}). \quad (11')$$

The factor 2 is essential and should be noted.

Such a potential function χ will only be suitable for our purposes if the sum of χ and Mc^2 represents the *total* energy of the moving particle in the "field of χ ." If we insert (11) and (11') in the identities (10) and (10') we find in fact that, following the motion,

$$\chi + mc^2\xi^{\frac{1}{2}} = \text{const.}$$

Thus χ represents the potential energy of the particle. Some simple manipulations can now be made to show that

$$\left(\mathbf{v} - \frac{\mathbf{P}}{t}\right) \cdot \left(-\frac{\partial \chi}{\partial \mathbf{P}}\right) = \frac{d}{dt} mc^2\xi^{\frac{1}{2}}. \quad (12)$$

This satisfactory equation shows that the force component $-\partial\chi/\partial\mathbf{P}$ increases the energy E at a rate equal to the work done in pushing the particle with the "ordinary" relative velocity $\mathbf{V} - \mathbf{P}/t$ of the particle relative to its surroundings in the substratum.

Finally, the equations of motion in a field of potential χ reduce, on inserting (11) and (11') in (9) and (9'), to

$$\frac{m\xi^{\frac{1}{2}}}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(\frac{\mathbf{V}}{Y^{\frac{1}{2}}}\right) = -m\xi^{\frac{1}{2}} \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y}\right) - \frac{\partial \chi}{\partial \mathbf{P}} + \frac{\mathbf{V}}{Y^{\frac{1}{2}}} \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (m\xi^{\frac{1}{2}}), \quad (13)$$

$$\frac{m\xi^{\frac{1}{2}}}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(\frac{c}{Y^{\frac{1}{2}}}\right) = -m\xi^{\frac{1}{2}} \left(ct - c \frac{Z}{Y}\right) + \frac{1}{c} \frac{\partial \chi}{\partial t} + \frac{c}{Y^{\frac{1}{2}}} \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (m\xi^{\frac{1}{2}}). \quad (13')$$

28. We must now remark that so far this dynamics is formal and conventional. It will only possess a content if we can deduce the form which a potential χ must have, and show deductively that the resulting motion reproduces some *natural* field of force. We could, indeed, have adopted other conventions for our definition of force; in that case χ would have appeared in a different rôle in the equations of motion, (13), (13'). But whatever conventions we adopt, we must be led to the same eventual motion when we apply our equations of motion to some actual situation; independently of our dynamical conventions, we must ultimately arrive at the motion that the abstract situation requires. Thus it does not really matter how we construct a dynamics; but having constructed it we must find actual forms for χ suitable to the abstract situations that are to have counterparts in the world of nature.

29. Now in a microscopic analysis only two types of "forces" are so far known in nature: gravitational and electromagnetic. Beginning with the first, as likely to be the more primitive, can we deduce, by further investigation of the properties of a substratum, what is the theoretical form of the most primitive type of "force," *i. e.* of χ , that presents itself? This is a problem of a higher order of difficulty than those so far considered.

It is well known that Einstein constructed a solution of the more primitive of these problems by *modifying* the properties of space. He modified these properties by imposing a relation between them and the description of the matter-in-motion occupying the space. The attempt

to go further, to impose still other restrictions on space, so as to embrace electromagnetic phenomena, is called the search for a "unified field theory." This search has been pursued with the utmost diligence, but has been generally adjudged to be unsuccessful. The reason may well be that it begins by seeking for "field equations," *i. e.* that it takes over the Faraday-Maxwell concept of a field (\mathbf{E} , \mathbf{H}), without further analysis, without recognizing that, as explained in Part I preceding, the Faraday-Maxwell theory is an induction from macroscopic experience and unsuitable in a micro-analysis of possibility.

We shall proceed differently. We shall solve the gravitational problem outright, in strictly relativistic fashion, in the flat private spaces of our fundamental observers. We shall find that this solution sets no bar to extension to electromagnetic forces, as Einstein's did, but rather suggests the way onward to them. Once the gravitational problem has been solved in *flat* space, the search for a "unified field theory" by further modifications of space is seen to be misguided. In our work, we shall make no hypotheses analogous to the setting up of "field equations." We shall rather explore the successive logical possibilities of the existence of forces. We shall find that gravitational forces exhaust the possibilities of forces represented by potentials χ , and that to find logically new types of forces we have to proceed to *super-potentials* ϕ , whose double partial differentiation yields something essentially different from gravitational forces. We eventually *deduce* the "field equations" which the vectors measuring the field satisfy; they appear as end-products, instead of as *ad hoc* suppositions.

30. To find the form of truly gravitational potentials χ we have to proceed indirectly. We have already seen that a substratum enjoys certain gravitational properties; we have seen that the interpretation of Newton's "constant" of gravitation in the present dynamics is $\gamma = c^3 t / M_0$. Now the proof of the equation of motion, (4), of a free particle in the presence of the substratum, not here so far reproduced, is somewhat complicated. It involves the consideration of the motion in time and space, of more recondite systems than the substratum, namely *statistical* systems of particles. To have the same description, in time- and space-co-ordinates, for each fundamental observer, these must have the spatio-velocity distribution law

$$N(\mathbf{P}, t, \mathbf{V}) dx dy dz du dv dw = \frac{\psi(\xi)}{c^6 X^{3/2} Y^{5/2}} dx dy dz du dv dw, \quad (14)$$

where $\psi(\xi)$ is an arbitrary function of ξ . Whatever function is chosen for $\psi(\xi)$, such a system enjoys the same property as a substratum, namely, that as observed by all fundamental particles $\mathbf{V} = \mathbf{P}/t$, the description in time, space and velocity is the same. Thus if O , O' are any two *fundamental* particles, defining, as usual, fundamental frames of reference $\mathbf{V} = \mathbf{P}/t$, we have both $O \equiv O'$ and $O \equiv O'$, as for the substratum, but now a statistical swarm, unlike the substratum, contains particles other than the fundamental particles.

The equation of motion of any member of the statistical swarm can be shown to be of the form

$$\mathbf{g} \equiv \frac{d\mathbf{V}}{dt} = \frac{Y}{X} (\mathbf{P} - \mathbf{V}t) G(\xi). \quad (15)$$

This reduces to the equation of motion of a free particle in the presence of the substratum alone, namely (4), on putting $G(\xi) \equiv -1$. Clearly the more general form (15) is due to the presence of the statistical swarm; the departure of the acceleration $d\mathbf{V}/dt$ from that characteristic of free motion in the presence of the substratum alone must afford a measure of the gravitational effect of the swarm, and so contains the key to a deductive approach to gravitation. The problem is now, can we determine $G(\xi)$, *i. e.* can we link $G(\xi)$ with $\psi(\xi)$, the function measuring the *population* of the swarm?

Now there is one condition we have not yet availed ourselves of: the condition that if every particle of the swarm undergoes an acceleration of the form (15), the swarm continues to be described by the same spatio-velocity distribution function $N(\mathbf{P}, t, \mathbf{V})$. (Clearly collisions are irrelevant in the present discussion: the particles may be considered as point-singularities of no finite size.) This condition is expressed by the generalized Boltzmann equation of gas-kinetic theory*,

$$\frac{\partial N}{\partial t} + \mathbf{V} \cdot \frac{\partial N}{\partial \mathbf{P}} + \mathbf{g} \cdot \frac{\partial N}{\partial \mathbf{V}} + N \left(\frac{\partial}{\partial \mathbf{V}} \cdot \mathbf{g} \right) = 0. \quad (16)$$

Hitherto in classical mathematical physics Boltzmann's equation has always been used in endeavours to determine N , the distribution function, given \mathbf{g} the acceleration. The novelty of our approach consists in using it to determine the acceleration \mathbf{g} , given N . The distribution function N we know already, from the general arguments (not here given) leading to (14). Thus we are in the unprecedented position of knowing the distribution N before we start. Hence we are in a position to *deduce* \mathbf{g} . It is precisely this step which is the novel feature of our analysis, the step which makes possible the deductive approach to problems which in Galileo's laws were studied by experiment, in Newton's by the assumption of physical law.

When we insert (14) and (15) in (16), we obtain a first-order ordinary differential equation in the variable ξ , which can be integrated and gives as the relation between $\psi(\xi)$ and $G(\xi)$

$$G(\xi) = -1 - \frac{C}{(\xi - 1)^{3/2} \psi(\xi)}, \quad (17)$$

C being the constant of integration. The appearance of a constant of integration in this relation is fundamental, and foreshadows the necessity for a number representing the *gravitational mass* of a particle. In particular, if we take the masses of the particles of the statistical swarm

* For a proof of this relation, and its integration, see the author's 'Relativity, Gravitation and World-Structure' (Oxford, 1935), §§ 207-217.

equal to zero, thus letting the swarm vanish, we get $G(\xi) = -1$, and we recover the equation of motion (4) of a free particle in the presence of the substratum alone.

Relations (14), (15), (17) embody the gravitational phenomenon, for they link density of population, defined through N or ψ , with acceleration g , defined by $G(\xi)$. Relation (17), derived purely deductively, must therefore contain within itself the seeds of a theory of gravitation.

31. Near a fundamental particle, $\xi \sim 1$. Moreover,

$$\xi - 1 = \frac{Z^2}{XY} - 1 = \frac{(\mathbf{P} - \mathbf{V}t)^2 - c^{-2}(\mathbf{P} \wedge \mathbf{V})^2}{c^2(t^2 - \mathbf{P}^2/c^2)(1 - \mathbf{V}^2/c^2)}, \quad \dots \quad (18)$$

so that near a fundamental particle (17) is approximately

$$G(\xi) = -1 - \frac{c^3 t^3}{|\mathbf{P} - \mathbf{V}t|^3} \frac{C}{\psi(1)}, \quad \dots \quad (19)$$

and thus the acceleration in this vicinity is given, by (15) and (19), by

$$g = \frac{d\mathbf{V}}{dt} \sim -\frac{\mathbf{P} - \mathbf{V}t}{t^2} - \frac{c^3 t}{M_0} \frac{M_0 C}{\psi(1)} \frac{\mathbf{P} - \mathbf{V}t}{|\mathbf{P} - \mathbf{V}t|^3}, \quad \dots \quad (20)$$

The first term here represents, as before, the pull of the whole substratum towards the (to P) apparent centre Vt . And then the second term represents an *inverse square attraction* towards the same centre. The constant of integration C plays the part of the gravitational mass associated with the singularity at $\xi = 1$, and, as before, the "constant" of gravitation is represented by $c^3 t/M_0$.

Thus two things, unforeseen, have emerged by purely deductive processes: (1) an inverse square attraction associated with a point singularity, and (2) a number which can be identified as gravitational mass.

32. If we keep to the exact equations instead of to approximations valid near $\xi \sim 1$, we find that the second term in the expression for the acceleration corresponds to a potential χ , in the precise sense defined above, given by

$$\chi = -\frac{m_1 m_2}{M_0} \frac{c^2 \xi^{\frac{1}{2}}}{(\xi - 1)^{\frac{1}{2}}}, \quad \dots \quad (21)$$

where now m_2 denotes the mass of the attracting condensation, standing for $M_0 C/\psi(1)$. From this, by a procedure I need not here particularize, we are led to the general expression

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{\xi^{\frac{1}{2}}}{(\xi - 1)^{\frac{1}{2}}}, \quad \dots \quad (22)$$

where

$$\xi = X_{12}^2 / X_1 X_2$$

and

$$X_1 = t_1^2 - \mathbf{P}_1^2 / c^2, \quad X_2 = t_2^2 - \mathbf{P}_2^2 / c^2,$$

$$X_{12} = t_1 t_2 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2.$$

Here (\mathbf{P}_1, t_1) , (\mathbf{P}_2, t_2) describe two particles, of masses m_1 and m_2 respectively, of which m_1 is the test particle or *attracted* particle, m_2 the field-

originating or *attracting* particle. It will be seen that χ is symmetrical in the particles, so that we can interchange their rôles without altering χ . χ represents, in fact, the mutual energy of the two particles.

33. Formula (22) successfully embodies the inverse square law of gravitation in Lorentz-invariant form, as we shall now see. Written out explicitly it becomes

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{X_{12}}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} = -\frac{m_1 m_2 c^3}{M_0} \cdot \frac{t_1 t_2 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2}{[(t_1 \mathbf{P}_2 - t_2 \mathbf{P}_1)^2 - (\mathbf{P}_1 \wedge \mathbf{P}_2)^2 / c^2]^{\frac{1}{2}}}. \quad (23)$$

To see the meaning of this, consider its value when the observer is at the field-generating particle m_2 . Then $\mathbf{P}_2 = 0$, and we have

$$[\chi]_{\mathbf{P}_2=0} = -\frac{m_1 m_2 c^3 t_1}{M_0} \frac{1}{|1|}. \quad (24)$$

This is the standard form of the empirical gravitational potential with $\gamma = c^3 t_1 / M_0$. It is to be noted that it is formally independent of the epoch t_2 at the field-originating particle. Again, the "attraction" is given by

$$-\frac{\partial \chi}{\partial \mathbf{P}_1} = \frac{m_1 m_2}{M_0} \frac{X_2}{(X_{12}^2 - X_1 X_2)^{3/2}} (\mathbf{P}_2 X_1 - \mathbf{P}_1 X_2). \quad (25)$$

When the observer is at the field-originating particle and $\mathbf{P}_2 = 0$, this gives

$$\left[-\frac{\partial \chi}{\partial \mathbf{P}_1} \right]_{\mathbf{P}_2=0} = -\frac{m_1 m_2 c^3 t_1}{M_0} \frac{\mathbf{P}_1}{|\mathbf{P}_1|^3}, \quad (26)$$

which is the precise form of the inverse square law with $\gamma = c^3 t_1 / M_0$ as before. It is particularly to be noted that care must be taken to *specify the observer* before the rigorous inverse square law can be formulated. A further point to notice is that, according to (25), the attraction $-\partial \chi / \partial \mathbf{P}_1$ vanishes when $X_2 = 0$, i. e., for $|\mathbf{P}_2| = ct_2$, i. e., for an attracting particle on the confines of the substratum or moving swarm. One other property of (22) may be mentioned here, namely, that it satisfies the wave equations

$$\begin{aligned} \frac{\partial^2 \chi}{\partial x_1^2} + \frac{\partial^2 \chi}{\partial y_1^2} + \frac{\partial^2 \chi}{\partial z_1^2} - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t_1^2} &= 0, \\ \frac{\partial^2 \chi}{\partial x_2^2} + \frac{\partial^2 \chi}{\partial y_2^2} + \frac{\partial^2 \chi}{\partial z_2^2} - \frac{1}{c^2} \frac{\partial^2 \chi}{\partial t_2^2} &= 0. \end{aligned}$$

34. It is very remarkable that whilst (23) and (25) appear to depend explicitly on the values of t_1 and t_2 , the values of the potential χ and attraction $-\partial \chi / \partial \mathbf{P}_1$ given by (24) and (26) depend only on t_1 , the epoch at the *attracted* particle. Since χ is a scalar, and since it is independent of t_2 when O is chosen at \mathbf{P}_2 , its value must be independent of t_2 for all positions of O. This shows that there is no need of any linkage between t_2 and t_1 ; in other words, the gravitational potential is not propagated as a wave—it exhibits no retardation effects. The *explicit* appearance of t_2 in (23) and (23') arises from the fact that for a

given attracting particle, the measure of its spatial position \mathbf{P}_2 depends on the epoch t_2 and on the observer O observing it.

35. The expression of the inverse square law, or its potential $1/r$, in a form relativistically invariant under Lorentz transformation from any one fundamental observer to any equivalent observer, as expressed by (23) or (23'), is of course revolutionary. For it was the supposed impossibility of executing this programme that led Einstein to forego the methods of so-called "special" relativity and develop the totally distinct procedure involved in "general" relativity. Unable to make progress with the law of gravitation for *all* transformations of co-ordinates which left $ds^2 = -dx^2 - dy^2 - dz^2 + c^2 dt^2$ unaltered, Einstein altered the form of ds^2 , that is, forced a curvature upon space. We now see that this resort was unnecessary. To require that the law of gravitation should be of constant form for *all* pairs of observers who arrived at the same ds^2 was too stringent a condition. For *all* such pairs of observers do not stand in the same relation to the substratum. We have imposed the only condition that it is legitimate to impose, namely, that the law should be of the same form for all pairs of *equivalent* observers, namely, for the class of fundamental observers in the substratum. The result is (23) or (24), which give χ as a 4-scalar, *i. e.*, as the same number whatever observer O is chosen.

I have often been asked whether this solution of the problem of gravitation satisfies the observational tests of the "general" theory of Einstein. The answer is that until some worker shows how to introduce non-equivalent observers into the present theory, the famous experimental tests of the general theory cannot be interpreted. The measures constituting the results of any experiment depend on the observer making them. An observer on the earth is not equivalent to an observer at a hypothetical nebular nucleus, and his measures will differ precisely in terms in inverse powers of c from those of an observer equivalent to an observer coincident with the centre of our galaxy. An immense field of research presents itself in the search for the ways in which the form of χ we have found will alter for non-equivalent observers such as a terrestrial observer. This is not an imperfection in the theory of gravitation we have constructed. It is sufficiently remarkable that without specific physical appeals we have been led, deductively, to an account of gravitation which reproduces the broad outline of gravitational phenomena, namely, the existence of an inverse square law and the numerical value of γ , in a form which is invariant for observers strictly equivalent to one another.

36. The first term in (17), the term -1 , corresponds to the equation of motion (4) of a *free* particle. For a *free* particle in the presence of the substratum alone, $G(\xi) \equiv -1$. We have interpreted this equation of motion (4) as giving the gravitational effect of the substratum on a free particle, and departures from this motion as "due" to external force. In Newtonian theory a free particle moves in "empty" space with

uniform velocity, and "force" is what causes a departure from such *uniform* motion. Can we reconcile the two states of affairs?

In the present theory, so far, all co-ordinates assigned by an observer O—the epochs, distances, velocities and accelerations he assigns to moving particles—have been based on readings made on his clock, reading time t . But he is always entitled at any time (i) to re-graduate his clock to read some other time, (ii) to change the conventions by which he constructs his co-ordinates of events from his observations of those events, and lastly (iii) to change the conventions by which he constructs and adopts a "space" out of those co-ordinates. Changes of conventions by which co-ordinates are constructed out of observations are clearly of little theoretical importance. But a re-graduation of the clock used, with, of course, corresponding re-graduations of the clocks of all equivalent observers, radically affects the descriptions of phenomena, whilst leaving the phenomena themselves unaffected. "General" relativity is not accustomed to distinguish between these two kinds of transformations of co-ordinates; and further, "general" relativity never changes the space it adopts. In fact, "general" relativity fails to bring in the observer explicitly. Kinematic relativity is more flexible. It distinguishes between re-graduations of observers' clocks and transformations of co-ordinates which are merely mathematical combinations of previous co-ordinates; it brings in the observer inevitably (as one of a swarm of *equivalent* observers) and it permits the selection of an appropriate space.

37. It is now found that if each equivalent observer re-graduates his clock from reading time t to time τ , where

$$\tau = t_0 \log (t/t_0) + t_0, \quad . \quad . \quad . \quad . \quad . \quad (27)$$

then the description of the substratum and of phenomena occurring in it are radically altered, whilst if t_0 is chosen to be the present value of t , *i. e.* the present age of the universe on the t -scale, actual measures of cosmically nearby events are scarcely changed. It is supposed that the same conventions are adopted for constructing co-ordinates out of observations as before. If a light-signal despatched from O at epoch t_1 or τ_1 to a distant event returns at epoch t_2 or τ_2 , then the measures (l, t) of distance and epoch on the t -scale become (λ, τ) on the τ -scale, where, since

$$\begin{aligned} l &= \frac{1}{2}c(t_2 - t_1), & t &= \frac{1}{2}(t_2 + t_1), \\ \lambda &= \frac{1}{2}c(\tau_2 - \tau_1), & \tau &= \frac{1}{2}(\tau_2 + \tau_1), \\ \tau_1 &= t_0 \log (t_1/t_0) + t_0, \\ \tau_2 &= t_0 \log (t_2/t_0) + t_0, \end{aligned}$$

we have

$$t = t_0 e^{\frac{\tau - t_0}{t_0}} \cosh \frac{\lambda}{ct_0}, \quad \frac{l}{c} = t_0 e^{\frac{\tau - t_0}{t_0}} \sinh \frac{\lambda}{ct_0}. \quad . \quad . \quad . \quad . \quad (28)$$

For nearby events ($l \ll ct$), these become

$$\tau = t_0 \log (t/t_1) + t_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

$$l = (t/t_0)\lambda. \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

so that since t can in present experience differ only very slightly from t_0 , l and τ are locally nearly coincident with λ and t .

38. It can be shown that the substratum, under this re-graduation of clocks, becomes an *infinite homogeneous, relatively stationary* distribution of particles or frames of reference. This must be contrasted with its previous description, in t -measure, as a *finite, non-homogeneous, expanding* distribution. The boundary $v=ct$ of the expanding substratum is transformed to infinity. The epoch τ , since it is connected with the invariant X by

$$X^{\frac{1}{2}} = (t^2 - l^2/c^2)^{\frac{1}{2}} = t_0 e^{\frac{\tau - t_0}{t_0}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

is itself an invariant, and, unlike t , is now the same, for a given event, for all the now-stationary fundamental observers. Instead of the flat *private* spaces of the moving fundamental observers, it is now possible to choose a *public* space, common to all, though this is now of negative curvature. The new space-time, $d\sigma^2$, is conformal to the old, being given by

$$ds = d\sigma e^{\frac{\tau - t_0}{t_0}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where the new spatial metric $d\epsilon^2$ is given by

$$d\sigma^2 = d\tau^2 - d\epsilon^2/c^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (32')$$

$$\text{with} \quad d\epsilon^2 = d\lambda^2 + (ct_0)^2 \sinh^2 \frac{\lambda}{ct_0} (d\theta^2 + \sin^2 \theta d\phi^2). \quad . \quad (32'')$$

The particle-density in the substratum, instead of being given by (2), becomes $B/(ct_0)^3$.

39. We must distinguish between (27) and (29): (27) is a *re-graduation of clocks*, and in it t , τ refer to epochs at the observer, measured by his own clock; (29) is the consequent *transformation of epoch-co-ordinates* for events near the observer. For events in general, we have, of course, the exact formulæ of transformation (28).

40. The equation of motion of a free particle, (4), under this re-graduation of clocks, suffers a remarkable change. The term representing the gravitational effect of the substratum wholly disappears, leaving an equation representing *uniform* motion in the space of metric (32''). Einstein accustomed us to the motion of a *local* gravitational field being transformed away by a change of co-ordinates: the present transformation transforms away the complete gravitational field of the cosmos.

Near the observer, under transformation (29) and (30), the equation of motion (4) of a free particle becomes simply

$$\frac{d^2 \Pi}{d\tau^2} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

where Π is the τ -measure of the vector \mathbf{P} . Since (33) is the equation of motion of a free particle in Newtonian mechanics, we can identify τ as the Newtonian time-variable.

It is worth while examining the path to (33) in more detail. Let us begin with equation (9) in the approximate form it takes for small velocities ($|\mathbf{V}| \ll c$) near the observer. This is

$$m \frac{d\mathbf{V}}{dt} = -m \frac{\mathbf{P} - \mathbf{V}t}{t^2} + \mathbf{F}. \quad (34)$$

Now by (30)

$$\mathbf{V} = \frac{d\mathbf{P}}{dt} = \frac{t_0}{t} \frac{d}{d\tau} \left[\frac{t}{t_0} \Pi \right] = \frac{d\Pi}{d\tau} + \frac{\Pi}{t_0} = \frac{d\Pi}{d\tau} + \frac{\mathbf{P}}{t},$$

whence

$$\mathbf{V} - \frac{\mathbf{P}}{t} = \frac{d\Pi}{d\tau}. \quad (35)$$

Thus the new measure of velocity, $d\Pi/d\tau$, represents the velocity relative to the environment in the expanding substratum. Differentiating (35) again, we have

$$\frac{d\mathbf{V}}{dt} - \frac{\mathbf{V}}{t} + \frac{\mathbf{P}}{t^2} = \frac{t_0}{t} \frac{d^2\Pi}{d\tau^2},$$

whence (34) gives

$$m \frac{d^2\Pi}{d\tau^2} = \frac{t}{t_0} \mathbf{F} = \Phi, \quad (36)$$

say. This is the Newtonian equation of motion under force Φ . In ordinary experience, the value of Φ is indistinguishable from the value of \mathbf{F} .

41. The gravitational potential χ transforms into

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \coth \frac{\lambda_2 - \lambda_1}{ct_0}, \quad (37)$$

when the observer O is in line with the two particles of co-ordinates λ_1 and λ_2 . The constant of gravitation thus reduces to $\gamma_0 = c^3 t_0 / M_0$, and in the τ -scale is truly constant once t_0 is fixed. But its numerical value depends on t_0 , our present epoch.

42. It may be asked why we have arrived at the Newtonian equation (33) by such a lengthy path. Why did we not begin with the τ -scale, and attempt to derive (33) directly? The answer is that (33) is only a local approximation; and that the exact form of the equations of motion of a free particle in the presence of the stationary substratum, valid for all distances, involve mention of the "constant" t_0 . This follows since t_0 appears in the formula (32'') for the metric. We were led to begin with formula (4). Formula (4) is the only form of the equation of motion of a free particle which does not contain a constant akin to t_0 . This is because the expanding substratum is the only description of the substratum which is free from constants. Formula (3) contains no constants in its make-up, save the conventional c and B ; consequently the equation of motion of a free particle in its presence can contain no fresh constants.

The existence of the natural zero of time allows us to state epochs t , and so build up expressions for accelerations. Without an *absolute* epoch we could never approach the problem of ascertaining accelerations. Hence the difficulty of attempting to establish Newtonian dynamics outright in an *a priori* fashion; this would be tantamount to beginning with a stationary substratum without knowing that its description contained an epoch t_0 . It would, of course, be quite foreign to our procedure to assume the existence of a parameter t_0 ; and to demonstrate it involves beginning with the t -dynamics. The parameter t_0 is a very strange one. It is constant for the calculations made by any one observer at any one stage in his history. But having selected his value for t_0 to correspond to his present epoch, he will find that his dynamics slowly departs from the elementary Newtonian formulation as his epoch changes.

43. In particular, we notice that the constant of gravitation $\gamma_0 = c^3 t_0 / M_0$ is built up out of t_0 together with two conventional numbers: c , which fixes the relation of the length and time units, here employed, to those of ordinary experience; and M_0 , the conventional member assigned to the apparent mass of the universe in the t -scale, which fixes our unit of mass in ordinary experience. It is highly satisfactory that this elusive "constant," the constant of gravitation, constructs itself in our synthesis in so simple a fashion out of c , t_0 and M_0 . If we take

$$c = 3 \times 10^{10} \text{ cm. sec.}^{-1}, \quad t_0 = 2 \times 10^9 \text{ years} = 0.6 \times 10^{17} \text{ sec.}, \\ \gamma_0 = 6.66 \times 10^{-8},$$

we have

$$M_0 = \frac{c^3 t_0}{\gamma_0} = \frac{27 \times 10^{30} \times 0.6 \times 10^{17}}{6.66 \times 10^{-8}} = 2.5 \times 10^{55} \text{ grams.}$$

Since $M_0 = \rho_0 \frac{4}{3} \pi (ct_0)^3$, this is equivalent to assigning to the mean density ρ_0 of matter near ourselves the value

$$\rho_0 = 10^{-27} \text{ grams. cm.}^{-3}.$$

This is in agreement with the astronomical estimates, and corresponds to the mean density of the matter of our own galaxy, distributed through the space between the galaxies near ourselves. The "age" $t_0 = 2 \times 10^9$ years $= 0.6 \times 10^{17}$ sec. is evaluated from Hubble's velocity-distance law recession at the rate 500 km. sec⁻¹ per 10^6 parsecs, which gives, since 1 parsec $= 3 \times 10^{18}$ cm.,

$$t_0 = t = \frac{|\mathbf{P}|}{|\mathbf{V}|} = \frac{10^6 \times 3 \times 10^{18}}{500 \times 10^5} = 0.6 \times 10^{17} \text{ sec.}$$

(See Proc. Roy. Soc. A, cliv. p. 47 (March 1936), for a fuller discussion.)

44. Many other properties of this dynamics and theory of gravitation have been worked out elsewhere. In particular, angular momentum has some remarkable features. I shall, however, here conclude this account of kinematic relativity by stating the form which the exact equations of motion (9), (9') take in τ -measure for local phenomena when

no restriction is placed on the velocities. Let \mathbf{v} denote $d\mathbf{\Pi}/d\tau$, which is the velocity relative to the now stationary fundamental observers. Then the equations of motion are found to be

$$\frac{1}{(1-\mathbf{v}^2/c^2)^{\frac{1}{2}}} \frac{d}{d\tau} \left[\frac{m\mathbf{v}}{1-\mathbf{v}^2/c^2} \right] = \Phi, \quad \dots \dots \dots (38)$$

$$\frac{1}{(1-\mathbf{v}^2/c^2)^{\frac{1}{2}}} \frac{d}{d\tau} \left[\frac{mc}{1-\mathbf{v}^2/c^2} \right] = \Phi_{\tau}, \quad \dots \dots \dots (38')$$

This force (Φ , Φ_{τ}) if derivable from a potential, a scalar χ , is given by

$$\Phi = -\frac{\partial \chi}{\partial \mathbf{\Pi}} + 2 \frac{\mathbf{v}}{1-\mathbf{v}^2/c^2} \frac{d}{d\tau} \left[\frac{m}{(1-\mathbf{v}^2/c^2)^{\frac{1}{2}}} \right], \quad \dots \dots \dots (39)$$

$$\Phi_{\tau} = +\frac{1}{c} \frac{\partial \chi}{\partial \tau} + 2 \frac{c}{1-\mathbf{v}^2/c^2} \frac{d}{d\tau} \left[\frac{m}{(1-\mathbf{v}^2/c^2)^{\frac{1}{2}}} \right], \quad \dots \dots \dots (39')$$

and the above equations of motion then reduce to

$$\frac{1}{(1-\mathbf{v}^2/c^2)^{3/2}} m \frac{d\mathbf{v}}{d\tau} = -\frac{\partial \chi}{\partial \mathbf{\Pi}}, \quad \dots \dots \dots (40)$$

$$0 = \frac{1}{c} \frac{\partial \chi}{\partial \tau}, \quad \dots \dots \dots (40')$$

From the last we have

$$-\frac{d\chi}{d\tau} = +\mathbf{v} \cdot \left(-\frac{\partial \chi}{\partial \mathbf{\Pi}} \right) = \frac{d}{d\tau} \left[\frac{mc^2}{(1-\mathbf{v}^2/c^2)^{\frac{1}{2}}} \right] = \frac{dE}{d\tau} \quad \dots \dots \dots (41)$$

so that

$$E + \chi = \text{const.}, \quad \dots \dots \dots (42)$$

where

$$E = mc^2/(1-\mathbf{v}^2/c^2)^{\frac{1}{2}}. \quad \dots \dots \dots (43)$$

E is thus the kinetic energy, and is, in fact, the transform of $\xi^{\frac{1}{2}}$. These equations of motion are different in form from those of Einstein's "special" relativity mechanics, but his were derived on the view that *all* frames in uniform relative motion were equivalent. Ours have been derived under less restrictive conditions, as explained above. Further, Einstein's equations, relating to t -time in their Lorentz-covariant aspect, whilst relating to τ -time in their dynamical aspect, suffer from confusion. We shall see when we proceed to electrodynamics that it is essential to be consistent when motion occurs at high velocities.

XIII. *Vapour-pressure and Electrodynamic Volume of Suspensions.*

By H. Löwy, Dr.Phil. *

[Received September 4, 1940.]

IN 1870, Lord Kelvin investigated the relation between the pressure of saturated vapour and the curvature of the liquid surface. In the case of small spheres (for instance water-drops) the relation has, according

* Communicated by the Author.

to R. v. Helmholtz (1886) and L. J. R. Wilson (Phil. Trans. 1897, clxxxix, p. 265), the form :

$$\ln \frac{P}{P_0} = \frac{2Sm}{r\rho RT} \quad \dots \dots \dots (1)$$

P is the pressure of saturated vapour in contact with liquid spheres of radius r , P_0 the saturated vapour-pressure over a flat surface at the same temperature T . S is the surface-tension, m the molecular weight, ρ the density of the liquid, and $R=8.3 \cdot 10^7$, the gas constant.

The rain-drops in clouds, as well as the water-films on the sand-grains of desert dunes, constitute suspensions of electrically conductive particles in a dielectric medium. The *electrodynamic volume* (Phil. Mag. 1938, xxvi. p. 453) of such suspensions is, as I have shown (Phil. Mag., July 1940),

$$V=[\bar{\epsilon}], \quad \dots \dots \dots (2)$$

where $\bar{\epsilon}$ is the dielectric constant of the suspension and $[\bar{\epsilon}] = \frac{\bar{\epsilon}-1}{\bar{\epsilon}+2}$. In the case of water-films on spherical nuclei, the electrodynamic volume (per c.c.) is equal to the geometrical volume within the external film-surface, that is

$$V = \frac{4\pi}{3} r^3 N, \quad \dots \dots \dots (3)$$

where N is the number of particles in 1 c.c. Substituting the value of r from (3) to (1) we obtain a *relation between the vapour-pressure and the dielectric constant of suspensions*

$$\ln \frac{P}{P_0} = \frac{2Sm}{\rho RT} \sqrt[3]{\frac{4\pi N}{3 [\bar{\epsilon}]}} \quad \dots \dots \dots (4)$$

In the case of soil, it is convenient to introduce in (4), instead of N , the porosity p . Designating by a the radius of the sand-grains, we have

$$\frac{4\pi}{3} a^3 N = 1-p, \quad \dots \dots \dots (5)$$

and obtain thus

$$\ln \frac{P}{P_0} = \frac{2Sm}{\rho RT a} \sqrt[3]{\frac{1-p}{[\bar{\epsilon}]}} \quad \dots \dots \dots (6)$$

In this form, the equation directly reveals the range of its validity and the validity of Lord Kelvin's theory: If, by measurement of $\bar{\epsilon}$ and p , we find that

$$[\bar{\epsilon}] < 1-p, \quad \dots \dots \dots (7)$$

we must consider this as a criterion that formula (6) is no longer applicable. The inequality (7) indicates (Phil. Mag., Jan. 1943) that the deeper solidified water-layers of the film have been more and more exposed by evaporation, and that the hydrologic equation (2) has lost its validity. On the other hand, the quantities S , ρ , P_0 are no more constants during the evaporation in this advanced stage of desiccation and, therefore, also formula (1) is no longer applicable.

For normal water at 18°C . ($T=291$), the constants have the values :

$P_0=15.477$ mm. Hg, $S=73$ dyne/cm., $\rho=1$, $m=18$;

therefore, $\ln P/P_0=10^{-7}/r$. In figs. 1 and 2, the vapour-pressure P is given as a function of the water content (which is proportional to the film radius): in fig. 1 according to formula (1) for normal water at 18°C .; in fig. 2 according to soil measurements of F. W. Parker ('Soil Science,' 1921, p. 43).

Fig. 1.

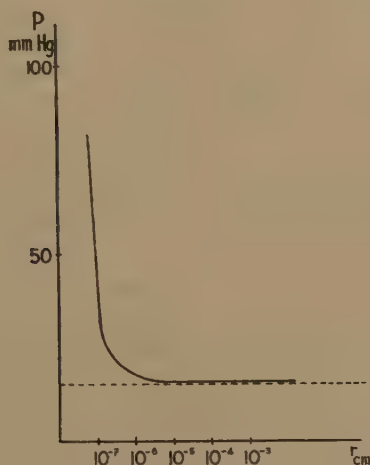
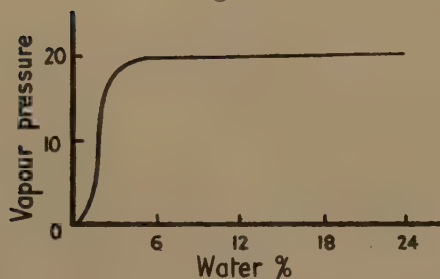


Fig. 2.



For the radius $r > 10^{-5}$, or water content greater than 4 per cent., both curves are asymptotes to the horizontal $P=P_0$. The pronounced change which occurs at 4 per cent. water content in Parker's curve, means, according to G. F. Bonyoncos ('Soil Science,' 1938, p. 45), that "the free water has ceased to operate and that the thin, discontinuous moisture-films have come into play." Formula (1) and fig. 1, however, clearly show that the horizontal part of Parker's curve (fig. 2) belongs to the film state, contrary to Bonyoncos' opinion. For radii of the film-spheres $r \geq 10^{-3}$ cm. (corresponding to fine sand), the curve (fig. 1),

according to formula (1), is nearly horizontal. The critical point in Parker's curve (fig. 2), corresponding to 4 per cent. water content, marks *not* the transition of free to film-water, but is situated within the film interval.

The question has a certain interest for the *biology of plants*, because the critical point in Parker's curve coincides with the *wilting point*. We can therefore conclude that *the desert plants utilize also film water* ('Bulletin de l'Institut d'Egypte,' 1940, xxii. p. 39).

There exists an essential difference between the non-horizontal branches of Kelvin's and Parker's curves. If, with diminishing drop radius, the vapour-pressure increases, the equilibrium of the drops is *labile*; in the contrary case *stable*. Let us, for instance, consider the case that under the real conditions, represented by Parker's curve (fig. 2), the equilibrium on the *non-horizontal* part of the curve between vapour-pressure P_1 and drop radius r_1 is disturbed by a small increase ΔP_1 of the vapour-pressure. This super-saturation will cause condensation, that is increase of the drop radius r , until the radius has reached its new equilibrium value. In an analogous way, a small decrease of P_1 will automatically produce a new equilibrium figure.

The partial pressure of vapour in the atmosphere being, in general, much smaller than the pressure of saturated vapour on a flat surface, we can conclude that the water-films in desert dunes are in *stable*, not in indifferent equilibrium, the latter corresponding to the horizontal branch of the curve ($P=P_0$). *In absence of rain and other influx, the water-content of the desert soil is determined by the partial pressure of the water-vapour in the atmosphere.*

An artificially dried sample of sand will, therefore, retain a certain quantity of film-water which changes with the vapour-pressure of the season. In his 'Contributions to the Geography of Egypt' (Cairo, 1939, p. 107), John Ball has observed that the water retained in dried residues of Nile silt amounts to 4.4 per cent. during the flood months, and to 2.2 per cent. during the other months. This result agrees qualitatively with Parker's curve (fig. 2), and the fact that, according to Sutton ('Meteorological Map of Egypt,' edited by Hurst), the vapour-pressure in Egypt during the flood months is twice as high as during the remainder of the year. Quantitatively, Ball's result, concerning the water content of dry sand, agrees very well with the value 2.3 per cent., which I have measured electrically in the Libyan Desert, on March 7, 1939, in the season of low vapour-pressure. As I have shown (Phil. Mag., July 1940) this value 2.3 per cent. approximately represents the quantity of non-solidified film-water. The same, no doubt, represents the two *limiting values of John Ball*.

When, with diminishing radius of the film-spheres, the quantities S , ρ , P_0 begin to change, the corresponding surface layers of the film are still liquid and electrically conductive. There exists, therefore, *beyond the range of Kelvin's theory*, an interval of r , in which *the hydrologic equation (2) has not yet lost its validity*. Combining the measurement of

vapour-pressure and dielectric constant, it will be possible to use the hydrologic equation for experimentally determining the relation between vapour-pressure and the curvature of the surface of water in the transition state between normal and solidified film-water.

Cairo.

June 18, 1940.

XIV. *The Boundary Couples in Thin Plates;*

By A. C. STEVENSON *.

[Received September 29, 1942.]

Introduction.

IN this paper the transformation of stress resultants and stress couples in a plate to curvilinear co-ordinates is carried out via the complex variables z , \bar{z} used by the writer in a previous paper ⁽¹⁾. This is instrumental in revealing an error occurring in Love's and Rayleigh's treatment of curvilinear co-ordinates in this connexion. Special attention is then paid to the simplifications introduced at a boundary by the boundary conditions, notably in the case of a clamped plate, where the stress resultants and stress couples are all expressible in terms of $\nabla^2 w$, w being the transverse displacement of the plate. This involves only one of the two complex potentials of the problem.

The paper goes on to find the complex potentials for the thin clamped plate of which the boundary curve is a regular curvilinear polygon of n "sides." The case $n=1$ reduces to that of the elliptic limaçon and contains that of the cardioid, both recently discussed by Sen ⁽²⁾ by another method and using different co-ordinates. As in the examples given by Sen, these results are interesting inasmuch as they can be expressed in finite terms. They include explicit results for the deflexion at the centre of the plate and for the boundary couples. The problem affords an example of the usefulness and simplicity of the complex potential method and of the boundary couple formulæ found in this paper.

1. *Transformation of Stress Resultants and Stress Couples.*

If s and n are arcs of orthogonal curves in the mid-plane of the plate, the stress couples $\tilde{n}s$, $\tilde{s}n$, $\tilde{s}s$ ($=-\tilde{n}n$) are best considered in the combinations

$$\Gamma' = \tilde{n}s + \tilde{s}n + 2i\tilde{s}s, \quad \Lambda' = \tilde{n}s - \tilde{s}n, \quad \dots \quad (1.1)$$

and are related to the cartesian stress couples $\tilde{x}y$, $\tilde{y}x$, $\tilde{y}y$ ($=-\tilde{x}x$) in their corresponding combinations

$$\Gamma = \tilde{x}y + \tilde{y}x + 2i\tilde{y}y, \quad \Lambda = \tilde{x}y - \tilde{y}x \quad \dots \quad (1.2)$$

* Communicated by the Author.

by the very simple equations of transformation (1.3),

$$\Gamma' = e^{-2i\alpha} \Gamma, \quad \Lambda' = \Lambda \quad . \quad . \quad . \quad . \quad . \quad (1.3)$$

where α is the angle made with the x -axis by the direction n at the point considered. The notation \widehat{pq} denotes the stress couple per unit area of cross-section of the plate perpendicular to the direction p , the axis of the couple being in the direction q .

Similarly, the mean stresses \widetilde{nz} , \widetilde{sz} and \widetilde{xz} , \widetilde{yz} across the thickness $2h$ of the plate, in the combinations

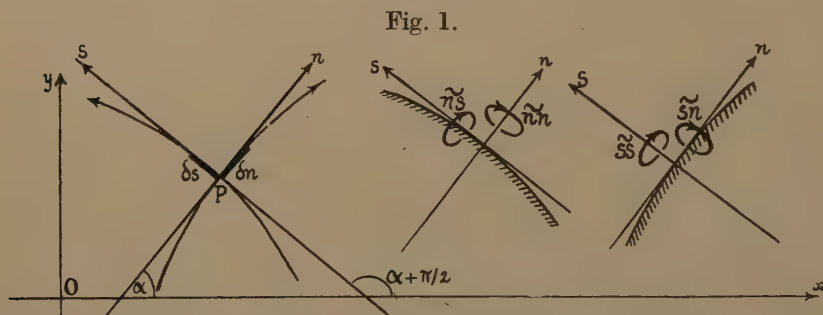
$$\Psi_0' = \widetilde{nz} + i\widetilde{sz}, \quad \Psi_0 = \widetilde{xz} + i\widetilde{yz}, \quad . \quad . \quad . \quad . \quad . \quad (1.4)$$

are related by the equation of transformation

$$\Psi_0' = e^{-i\alpha} \Psi_0 \quad . \quad . \quad . \quad . \quad . \quad (1.5)$$

and to the stress couples by the complex identical relation

$$\Psi_0 = \frac{\partial \Gamma}{\partial z} + \frac{\partial \Lambda}{\partial \bar{z}} \quad . \quad . \quad . \quad . \quad . \quad (1.6)$$



Scheme of curvilinear co-ordinates and stress couples.

For these and other results, quoted in this section without proof, the reader should consult the previous paper⁽¹⁾.

In the elementary theory of thin plates—and in certain cases in the theory of moderately thick plates—the stress couples are given in terms of the transverse displacement w by the equations

$$\Gamma = -4(1-\eta)P \frac{\partial^2 w}{\partial \bar{z}^2}, \quad \Lambda = -4(1+\eta)P \frac{\partial^2 w}{\partial z \partial \bar{z}}, \quad . \quad . \quad . \quad . \quad (1.7)$$

where η is Poisson's ratio and P is a plate modulus connected with the more usual flexural rigidity D of the plate by the relation

$$D = 2hP. \quad . \quad . \quad . \quad . \quad . \quad (1.8)$$

To carry out the transformation to the orthogonal arcs s , n and exhibit the stress resultants and stress couples in terms of differentiations intrinsic to these arcs, we have from the elementary arcs δs , δn of fig. 1, remembering the Argand diagram vector properties of complex numbers,

$$\frac{\partial z}{\partial n} = e^{i\alpha}, \quad \frac{\partial z}{\partial s} = e^{i(\alpha + \pi/2)} = ie^{i\alpha} = i \frac{\partial z}{\partial n}, \quad . \quad . \quad . \quad . \quad (1.9)$$

and the results
$$\frac{\partial w}{\partial n} = \frac{\partial w}{\partial z} \frac{\partial z}{\partial n} + \frac{\partial w}{\partial \bar{z}} \frac{\partial \bar{z}}{\partial n}, \quad \dots \quad (1.10)$$

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial z} \frac{\partial z}{\partial s} + \frac{\partial w}{\partial \bar{z}} \frac{\partial \bar{z}}{\partial s} = i \left\{ \frac{\partial w}{\partial z} \frac{\partial z}{\partial n} - \frac{\partial w}{\partial \bar{z}} \frac{\partial \bar{z}}{\partial n} \right\}, \quad \dots \quad (1.11)$$

on using (1.9).

From (1.4)–(1.9) we find

$$\Psi'_0 = 8iP \frac{\partial \bar{z}}{\partial s} \frac{\partial}{\partial \bar{z}} \left(\frac{\partial^2 w}{\partial z \partial \bar{z}} \right) = 2iP \frac{\partial \bar{z}}{\partial s} \frac{\partial}{\partial \bar{z}} \nabla^2 w,$$

since

$$\nabla^2 w = \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} = 4 \frac{\partial^2 w}{\partial z \partial \bar{z}},$$

and then separation of real and imaginary parts, using (1.10) and (1.11)

leads to
$$\bar{n}z = P \frac{\partial}{\partial n} \nabla^2 w, \quad \bar{s}z = P \frac{\partial}{\partial s} \nabla^2 w, \quad \dots \quad (1.12)$$

or the stress resultants throughout the plate are expressible in terms of $\nabla^2 w$. From (1.7) it is clear that the stress couple combination is also expressible throughout the plate in terms of $\nabla^2 w$.

Now from (1.9)

$$\frac{\partial^2 z}{\partial s^2} = -e^{i\alpha} \frac{\partial \alpha}{\partial s} = i \frac{\partial z}{\partial s} \frac{\partial \alpha}{\partial s} = -\frac{\partial z}{\partial n} \frac{\partial \alpha}{\partial s}, \quad \frac{\partial z}{\partial s} \frac{\partial \bar{z}}{\partial s} = 1, \quad \dots \quad (1.13)$$

and so, forming $\partial^2 w / \partial s^2$, we have

$$\frac{\partial^2 w}{\partial s^2} = \frac{\partial^2 w}{\partial s} \frac{\partial z}{\partial s} \frac{\partial s}{\partial s} + \frac{\partial w}{\partial z} \frac{\partial^2 z}{\partial s^2} + \frac{\partial^2 w}{\partial s} \frac{\partial \bar{z}}{\partial s} \frac{\partial s}{\partial s} + \frac{\partial w}{\partial \bar{z}} \frac{\partial^2 \bar{z}}{\partial s^2},$$

or, on using (1.11) and (1.13)

$$\frac{\partial^2 w}{\partial s^2} = \frac{\partial^2 w}{\partial z^2} \left(\frac{\partial z}{\partial s} \right)^2 + \frac{\partial^2 w}{\partial \bar{z}^2} \left(\frac{\partial \bar{z}}{\partial s} \right)^2 + 2 \frac{\partial^2 w}{\partial z \partial \bar{z}} - \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial s}.$$

Proceeding in this manner to form $\partial^2 w / \partial n^2$, $\partial^2 w / \partial n \partial s$, $\partial^2 w / \partial s \partial n$, we find

$$\frac{\partial^2 w}{\partial \bar{z}^2} \left(\frac{\partial \bar{z}}{\partial s} \right)^2 + \frac{\partial^2 w}{\partial z^2} \left(\frac{\partial z}{\partial s} \right)^2 = \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial s} + \frac{\partial^2 w}{\partial s^2} - \frac{1}{2} \nabla^2 w, \quad \dots \quad (1.14)$$

$$= \frac{\partial w}{\partial s} \frac{\partial \alpha}{\partial n} - \frac{\partial^2 w}{\partial n^2} + \frac{1}{2} \nabla^2 w, \quad \dots \quad (1.15)$$

$$i \left\{ \frac{\partial^2 w}{\partial \bar{z}^2} \left(\frac{\partial \bar{z}}{\partial s} \right)^2 - \frac{\partial^2 w}{\partial z^2} \left(\frac{\partial z}{\partial s} \right)^2 \right\} = \frac{\partial^2 w}{\partial n \partial s} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial n}, \quad \dots \quad (1.16)$$

$$= \frac{\partial^2 w}{\partial s \partial n} - \frac{\partial w}{\partial s} \frac{\partial \alpha}{\partial s}. \quad \dots \quad (1.17)$$

In particular we see from these that

$$\nabla^2 w = \frac{\partial^2 w}{\partial n^2} + \frac{\partial^2 w}{\partial s^2} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial s} - \frac{\partial w}{\partial s} \frac{\partial \alpha}{\partial n}, \quad \dots \quad (1.18)$$

$$\frac{\partial^2 w}{\partial n \partial s} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial n} = \frac{\partial^2 w}{\partial s \partial n} - \frac{\partial w}{\partial s} \frac{\partial \alpha}{\partial s}. \quad \dots \quad (1.19)$$

Now from (1.3) and (1.9)

$$\Gamma' + \bar{\Gamma}' = 4(1 - \eta)P \left\{ \frac{\partial^2 w}{\partial \bar{z}^2} \left(\frac{\partial \bar{z}}{\partial s} \right)^2 + \frac{\partial^2 w}{\partial z^2} \left(\frac{\partial z}{\partial s} \right)^2 \right\},$$

$$\Gamma' - \bar{\Gamma}' = 4(1 - \eta)P \left\{ \frac{\partial^2 w}{\partial \bar{z}^2} \left(\frac{\partial \bar{z}}{\partial s} \right)^2 - \frac{\partial^2 w}{\partial z^2} \left(\frac{\partial z}{\partial s} \right)^2 \right\},$$

and since the stress couples $\tilde{n}s$, $\tilde{s}n$, $\tilde{n}n$, $\tilde{s}s$ can be separated out as

$$4\tilde{n}s = \Gamma' + \bar{\Gamma}' + 2\Lambda', \quad 4\tilde{s}n = \Gamma' + \bar{\Gamma}' - 2\Lambda', \quad 4\tilde{n}n = -4\tilde{s}s = i(\Gamma' - \bar{\Gamma}'),$$

we find, using (1.14)–(1.17) and (1.7),

$$\tilde{n}s = (1 - \eta)P \left\{ \frac{\partial^2 w}{\partial s^2} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial s} \right\} - P \nabla^2 w. \quad . \quad . \quad . \quad (1.20)$$

$$\tilde{s}n = (1 - \eta)P \left\{ \frac{\partial^2 w}{\partial s^2} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial s} \right\} + \eta P \nabla^2 w, \quad . \quad . \quad . \quad (1.21)$$

$$\tilde{n}n = -\tilde{s}s = (1 - \eta)P \left\{ \frac{\partial^2 w}{\partial n \partial s} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial n} \right\} \quad . \quad . \quad . \quad (1.22)$$

$$= (1 - \eta)P \left\{ \frac{\partial^2 w}{\partial s \partial n} - \frac{\partial w}{\partial s} \frac{\partial \alpha}{\partial s} \right\} \quad . \quad . \quad . \quad (1.23)$$

An alternative form for $\tilde{n}s$, using (1.18), is

$$\tilde{n}s = -P \left\{ \frac{\partial^2 w}{\partial n^2} - \frac{\partial w}{\partial s} \frac{\partial \alpha}{\partial n} \right\} - \eta P \left\{ \frac{\partial^2 w}{\partial s^2} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial s} \right\} \quad . \quad . \quad . \quad (1.24)$$

Now (1.18), (1.19) are well-known two-dimensional curvilinear results, agreeing for example with equivalent analysis in Coker and Filon's 'Photo-Elasticity,' 1931, p. 162, where the operand w is replaced by a stress function χ , yet our results (1.22) and (1.24), for example, do not agree with those given by Love in his 'Mathematical Theory of Elasticity,' 1927, p. 465, for G, H (our $\tilde{n}s$, $\tilde{n}n$), unless we set $\partial\alpha/\partial n = 0$, which is clearly not the case in general. Actually Love mis-states that $\partial\theta/\partial\nu$ (our $\partial\alpha/\partial n$) = 0. He refers the reader to Rayleigh's 'Theory of Sound,' vol. i. §216, p. 357, which also gives results only in agreement with ours if we set $\partial\alpha/\partial n = 0$. In Rayleigh's case the error would seem to occur through his writing

$$\nabla^2 w = \frac{d^2 w}{dn^2} + \frac{d^2 w}{d\sigma^2},$$

where σ is a fixed axis coinciding with the tangent s at the point under consideration, and failing to notice that, equally well, he should also have taken, instead of n , a fixed axis ν coinciding with the normal n at the point under consideration. Since neither Love nor Rayleigh used their formulæ for other than cartesian or polar co-ordinates, in which cases $\partial\alpha/\partial n$ is strictly zero, there are no consequential errors to trace.

2. The Boundary Couples.

Consider the simplifications introduced at an actual boundary into the formulæ (1.20)–(1.24), which are valid throughout the plate, by

This gives an alternative form for the peripheral couple $\widetilde{s}n_0$,

$$\widetilde{s}n_0 = -(1-\eta^2)P \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial s} \quad (2.11)$$

Hence $\widetilde{s}n_0$ vanishes along a rectilinear simply supported boundary, as also does A , in which case the boundary conditions are

$$w=0, \quad \nabla^2 w=0 \quad (2.12)$$

(cp. Timoshenko 'Theory of Plates and Shells,' (p. 100, 1940).

Finally, consider the case of a *free* boundary, for which the Kirchhoff boundary conditions are

$$\widetilde{s}n_0=0, \quad \widetilde{n}z_0 - \frac{\partial \widetilde{n}n}{\partial s}=0. \quad (2.13)$$

From the first of these, $\widetilde{s}n_0$ is again given by (2.8), and in terms of w the boundary conditions become

$$\nabla^2 w = (1-\eta) \left\{ \frac{\partial^2 w}{\partial s^2} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial s} \right\}, \quad (2.14)$$

$$\frac{\partial}{\partial n} \nabla^2 w = (1-\eta) \frac{\partial}{\partial s} \left\{ \frac{\partial^2 w}{\partial n \partial s} + \frac{\partial w}{\partial n} \frac{\partial \alpha}{\partial n} \right\}. \quad (2.15)$$

It is worth pointing out that even in those cases of the "semi-inverse" method ⁽¹⁾ in the theory of moderately thick plates, where the equations (1.7) are not exactly satisfied, some simplification still results from forming the boundary stress couples after using the boundary conditions as above. Thus, for example, in the case of the clamped boundary, whilst we do not quite retain the simple elegance of the results (2.3), we still obtain a clamping couple $\widetilde{n}z_0$ in terms of one complex potential only.

3. Curvature of Orthogonal Curvilinear Curves.

In practice we use families of curves given by $\xi=\text{const.}$, $\eta=\text{const.}$, where

$$z=z(\zeta), \quad \zeta=\xi+i\eta \quad (3.1)$$

(there will be no occasion to confuse the η with Poisson's ratio). For then, writing

$$\delta z = z'(\zeta) \delta \zeta = J e^{i\alpha} \delta \zeta \quad \text{or} \quad z'(\zeta) = J e^{i\alpha}, \quad (3.2)$$

where J and α are real, we have, taking the curve $\eta=\text{const.}$ as the arc n of the previous sections,

$$\delta z_1 = J \delta \xi e^{i\alpha} = \delta n e^{i\alpha}, \quad \text{or} \quad \frac{\partial z}{\partial n} = e^{i\alpha}, \quad (3.3)$$

which identifies α with our previous α . Similarly, along the curve $\xi=\text{const.}$, we have

$$\delta z_2 = J d\eta e^{i(\alpha+\pi/2)} = \delta s e^{i(\alpha+\pi/2)}, \quad \text{or} \quad \frac{\partial z}{\partial s} = i \frac{\partial z}{\partial n}, \quad (3.4)$$

and since clearly $\delta z_1, \delta z_2$ represent perpendicular increments on an Argand diagram, we see that the curve $\xi = \text{const.}$ is the arc s of our previous work.

To find the curvatures $\partial\alpha/\partial s, \partial\alpha/\partial n$ of the arcs at the point considered, we have

$$\frac{\partial\alpha}{\partial s} = \frac{\partial\alpha}{\partial z} \frac{\partial z}{\partial s} + \frac{\partial\alpha}{\partial \bar{z}} \frac{\partial \bar{z}}{\partial s}, \quad \frac{\partial\alpha}{\partial n} = \frac{\partial\alpha}{\partial z} \frac{\partial z}{\partial n} + \frac{\partial\alpha}{\partial \bar{z}} \frac{\partial \bar{z}}{\partial n},$$

and using (3.4) this gives

$$\frac{\partial\alpha}{\partial s} + i \frac{\partial\alpha}{\partial n} = 2 \frac{\partial\alpha}{\partial z} \frac{\partial z}{\partial s} \quad \dots \quad (3.5)$$

But, from (3.2),

$$\log J + i\alpha = \log z'(\zeta) = \text{function of } z \text{ only,}$$

and so

$$i \frac{\partial\alpha}{\partial \bar{z}} = -\frac{1}{J} \frac{\partial J}{\partial \bar{z}}, \quad \text{or} \quad \frac{\partial\alpha}{\partial z} = -\frac{i}{J} \frac{\partial J}{\partial z}.$$

Also, using (3.2)-(3.4),

$$\frac{\partial z}{\partial s} = ie^{i\alpha} = \frac{iz'(\zeta)}{J},$$

$$\text{so that (3.5) becomes} \quad \frac{\partial\alpha}{\partial s} + i \frac{\partial\alpha}{\partial n} = \frac{2}{J^2} \frac{\partial J}{\partial \zeta} = \frac{1}{J^3} \frac{\partial J^2}{\partial \zeta} \quad \dots \quad (3.6)$$

And since

$$J^2 = z'(\zeta)\bar{z}'(\bar{\zeta}),$$

$$\frac{\partial\alpha}{\partial s} + i \frac{\partial\alpha}{\partial n} = \frac{z''(\zeta)\bar{z}'(\bar{\zeta})}{\{z'(\zeta)\bar{z}'(\bar{\zeta})\}^{3/2}}, \quad \dots \quad (3.7)$$

giving a simple formula for the derivation of the curvatures, when required.

4. Clamped Circular and Elliptic Plates.

The solution for the clamped circular plate of radius a is well known

$$\text{to be} \quad w = -W \left(1 - \frac{r^2}{a^2}\right)^2, \quad W = pa^4/64D, \quad \dots \quad (4.1)$$

since this satisfies $w=0, \partial w/\partial r=0$ round $r=a$, and $\nabla^4 w = -p/D$ throughout the plate, W being the actual numerical deflexion at the centre of the plate. Since $\nabla^2 w = -8W/a^2$ round $r=a$, (2.3) gives for the clamping couple $\widetilde{r\theta}_0$,

$$\widetilde{r\theta}_0 = 8PW/a^2 = pa^3/16h. \quad \dots \quad (4.2)$$

Similarly, Bryan's solution for the clamped elliptic plate of boundary

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

$$\text{is} \quad w = -W \left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)^2, \quad W = a^4 b^4 p/8D(3a^4 + 3b^4 + 2a^2 b^2), \quad (4.3)$$

since this satisfies $w=0, \partial w/\partial n=0$ round the elliptic boundary, and $\nabla^4 w = -p/D$ throughout the plate. If ϖ is the perpendicular from the centre upon the tangent to the ellipse at the point of the boundary under consideration, we have

$$\nabla^2 w = -8W \left(\frac{x^2}{a^4} + \frac{y^2}{b^4}\right) = -\frac{8W}{\varpi^2},$$

and (2.3) gives for the clamping couple \widetilde{ns}_0

$$\widetilde{ns}_0 = 8PW/\pi^2. \quad (4.4)$$

Hence the maximum and minimum values of the clamping couple are at the ends of the minor and major axes respectively, since $b \leq \pi \leq a$ (cp. Timoshenko 'Theory of Plates and Shells,' p. 292 (1940), where, however, (4.4) is not given explicitly).

5. The Clamped Curvilinear Polygonal Plate.

Let the boundary curve C_0 in the z -plane be such as to map into the unit circle C in the σ -plane under the conformal transformation

$$z = c\sigma(1 + \lambda\sigma^n), \quad (5.1)$$

where we take c and λ to be real and positive constants, and, in order that the transformation shall be conformal at all points within the boundary, $z'(\sigma)$ must not vanish or become infinite at points within C_0 , and we therefore have the restriction

$$0 \leq \lambda(n+1) \leq 1. \quad (5.2)$$

Then, taking curvilinear co-ordinates ξ, η such that

$$\sigma = e^{\xi}, \quad \zeta = \xi + i\eta, \quad (5.3)$$

the boundary C_0 will be the curve $\xi=0$, and, along this curve, putting $z = re^{i\theta}$, we have from (5.1) and (5.3)

$$r^2 = z\bar{z} = c^2(1 + \lambda^2 + 2\lambda \cos n\eta), \quad (5.4)$$

$$\tan \theta = \frac{\sin \eta + \lambda \sin (n+1)\eta}{\cos \eta + \lambda \cos (n+1)\eta}. \quad (5.5)$$

Hence the curve C_0 is such that when r is stationary, $\sin n\eta=0$ and

$$\eta = \theta = r\pi/n, \quad r=0, 1, 2, \dots, 2n-1.$$

It therefore appears that C_0 is, in general, a certain regular curvilinear polygon of n "sides" and n rounded vertices, and such that

$$n/(n+1) \leq 1 - \lambda \leq r \leq 1 + \lambda \leq (n+2)/(n+1).$$

The rounded vertices or indentations of the curve occur where r is a minimum, and become angular points (cusps) when the critical points of the conformal transformation, given by $z'(\sigma)=0$, fall upon the boundary C_0 . In the case $n=1$, a simple change of origin given by $z' = z + \lambda c$ makes

$$z' = c\sigma[1 + \lambda(\sigma + \sigma^{-1})] = ce^{in}(1 + 2\lambda \cos \eta), \quad 0 \leq 2\lambda \leq 1 \quad (5.6)$$

so that, in polar co-ordinates, given here by $z' = re^{i\theta}$, we recognize the boundary C_0 as the elliptic limaçon

$$r = a + b \cos \theta, \quad b \leq a, \quad (a=c, b/a=2\lambda) \quad (5.7)$$

passing from the circle $r=a$, when $\lambda=0$, to the cardioid when $b=a$, i. e. $\lambda=1/2$.

If we write $\Omega_0(z)=\Omega(\sigma)$, $\omega_0(z)=\omega(\sigma)$, then when the boundary C_0 is clamped, the boundary conditions (2.6) give, using (2.4) and the equation of the boundary $\sigma\bar{\sigma}=1$,

$$\bar{z}(\sigma^{-1})\Omega(\sigma)+z(\sigma)\bar{\Omega}(\sigma^{-1})+\omega(\sigma)+\bar{\omega}(\sigma^{-1})+z^2(\sigma)\bar{z}^2(\sigma^{-1})=0, \quad (5.8)$$

$$\bar{z}(\sigma^{-1})\Omega'(\sigma)+z'(\sigma)\bar{\Omega}(\sigma^{-1})+\omega'(\sigma)+2\bar{z}^2(\sigma^{-1})z(\sigma)z'(\sigma)=0. \quad (5.9)$$

These might very well be solved by Muschelisvili's direct method⁽³⁾, but on the whole it is easier to use the usual tentative method after an inspection of the explicit boundary conditions. From (5.1)

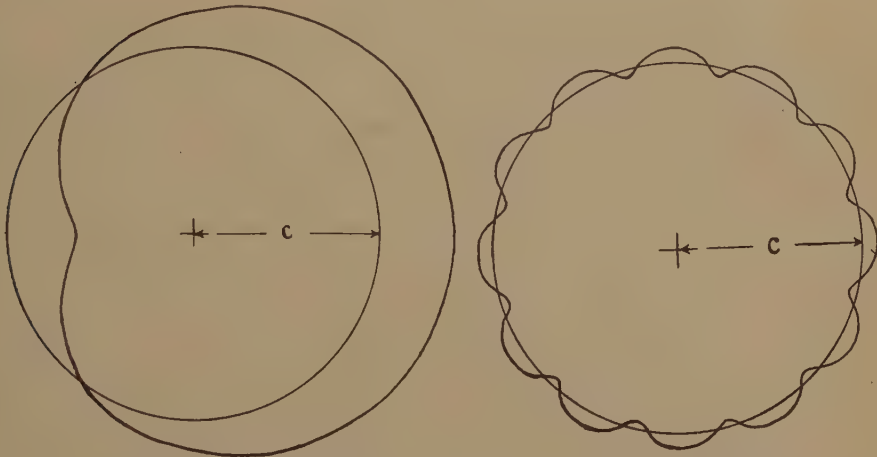
$$2\bar{z}^2(\sigma^{-1})z(\sigma)z'(\sigma)=2c^4\sigma^{-1}\{\alpha+\beta\sigma^n+\gamma\sigma^{-n}+\delta\sigma^{2n}+\epsilon\sigma^{-2n}\},$$

where

$$\alpha=1+2(n+2)\lambda^2+(n+1)\lambda^4, \quad \beta=\lambda[n+2+2(n+1)\lambda^2],$$

$$\gamma=\lambda[2+\lambda^2(n+2)], \quad \delta=(n+1)\lambda^2, \quad \epsilon=\lambda^2, \quad \dots \dots \dots (5.10)$$

Fig. 2.



(i) $n=1$, $\lambda=3/8(<1/2)$.

(ii) $n=12$, $\lambda=1/16(<1/13)$.

Examples of the boundary curve C_0 .

so that, if we take

$$\Omega(\sigma)=c^3(A\sigma+B\sigma^{n+1}+C\sigma^{2n+1}), \quad \omega(\sigma)=c^4(E+F\sigma^n+G\sigma^{2n}), \quad (5.11)$$

where A, B, C, E, F, G are real constants, then the boundary condition (5.9) can be satisfied by equating the coefficients of $\sigma^0, \sigma^n, \sigma^{-n}, \sigma^{2n}, \sigma^{-2n}$ separately to zero. This gives five equations for A, B, C, F, G :

$$A+(n+1)\lambda B+\alpha=0,$$

$$(n+1)(\lambda A+B)+(2n+1)\lambda C+nF+2\beta=0,$$

$$\lambda A+B+(n+1)\lambda C+2\gamma=0,$$

$$(2n+1)C+2nG+2\delta=0, \quad C+2\epsilon=0,$$

$$\text{whence } A=-\frac{1-2n\lambda^2-(n+1)\lambda^4}{1-\lambda^2(n+1)}, \quad B=-\frac{\lambda[3-2(n+1)\lambda^2-(n+1)\lambda^4]}{1-\lambda^2(n+1)}, \quad \left. \begin{aligned} C &=-2\lambda^2, & F &=2\lambda(1+\lambda^2), & G &=\lambda^2 \end{aligned} \right\} \quad (5.12)$$

Treating (5.8) in the same way, we obtain the equations

$$\begin{aligned} G + C + \lambda^2 &= 0, \\ \lambda A + B + \lambda C + F + 2\lambda(1 + \lambda^2) &= 0, \\ 2(A + \lambda B + E) + (1 + \lambda^2)^2 + 2\lambda^2 &= 0, \end{aligned}$$

of which the first two are consistent with (5.12), and the third gives the remaining constant E as

$$E = \frac{1}{2} \frac{1 - 3(n-1)\lambda^2 - (2n+3)\lambda^4 - (n+1)\lambda^6}{1 - (n+1)\lambda^2}. \quad (5.13)$$

In particular (2.4) and (5.13) give for the deflexion W at the origin $z=0$ ($w=-W$),

$$W = \frac{pc^4}{64D} \frac{1 - 3(n-1)\lambda^2 - (2n+3)\lambda^4 - (n+1)\lambda^6}{1 - (n+1)\lambda^2}. \quad (5.14)$$

For $\lambda=0$, the plate becomes a circle of radius c and the result (5.14) gives that of (4.1). As a further numerical check, the formula due to Sen [see (1.14) of his paper⁽²⁾] in the case of the cardioid gives the deflexion at our origin, on putting his $X=1/2$, as $W=21pa^4/1024 D$ in our notation. On putting $n=1$, $\lambda=1/2$ in (5.14) we have the same result.

From (2.5), the clamping couple $\xi\eta_0$ is given by

$$\xi\eta_0 = \frac{p}{32h} \left\{ \frac{\Omega'(\sigma)}{z'(\sigma)} + \frac{\bar{\Omega}'(\sigma^{-1})}{\bar{z}'(\sigma^{-1})} + 4z(\sigma)\bar{z}(\sigma^{-1}) \right\}. \quad (5.15)$$

Using (5.11)–(5.13) we obtain, after some reduction, the clamping couple as

$$\xi\eta_0 = \frac{pc^2}{32h} (A'\varpi^2 + B' + C'\varpi^{-2}), \quad (5.16)$$

$$\begin{aligned} \text{where } \varpi^2 &= 1 + \lambda^2(n+1)^2 + 2\lambda(n+1) \cos n\eta, \quad A' = 2/(n+1)^2, \\ \text{and } B' &= 4n/(n+1)^2, \quad C' = \frac{2n[n + \lambda^2(n+1)][1 - \lambda^2(n+1)^2]}{(n+1)^2[1 - \lambda^2(n+1)]} \end{aligned} \quad (5.17)$$

Now the boundary reduces to a circle if $\lambda=0$, which makes $\varpi=1$, and $A'+B'+C'=2$, so that $\xi\eta_0=pc^2/16h$, agreeing with the result (4.2) for the circular plate, as it should.

Notice that since $\Omega(\sigma)$ and $\omega(\sigma)$ are holomorphic throughout the plate, and $z'(\sigma)$ does not vanish at any point in the plate by the restriction (5.2), the stresses and couples are physically admissible throughout the plate.

References.

- (1) Stevenson, Phil. Mag. xxxiii. p. 639 (1942).
- (2) Sen, Phil. Mag. xxxiii. p. 294 (1942).
- (3) Muschelishvili, Z. A. M. M. xiii. p. 264 (1933).

XV. *The Torsion of a Fluted Column.*

By A. C. STEVENSON †.

[Received September 29, 1942.]

THIS note considers briefly the torsion of a cylinder of simply-connected cross-section by an extension of a method used elsewhere⁽¹⁾ by the writer, and deduces the results recently obtained by Morris⁽²⁾. The method is illustrated by applying it to the problem of a grooved or fluted column with a cross-section of which the boundary is the regular curvilinear polygon of n "sides" illustrated in the previous paper⁽³⁾.

1. *Harmonic Boundary Problems.*

The type of harmonic boundary problem envisaged here concerns a domain S_1 in the z -plane, containing the origin, and bounded by a simple closed curve C_1 , and consists of finding a complex potential $\omega_1(z) = \phi + i\psi$, holomorphic (regular, analytic) within S_1 , and such that ψ has prescribed values along C_1 . The method of solution adopted here is to map the domain S_1 on the interior S of unit circle C in a σ -plane by a conformal transformation

$$z = z(\sigma) = \sum_0^{\infty} a_r \sigma^r, \quad z'(\sigma) \neq 0 \text{ or } \infty \text{ within } S, \quad . \quad . \quad (1.1)$$

so that the problem is equivalent to finding a complex potential

$$\omega(\sigma) = \omega_1[z(\sigma)] = \sum_0^{\infty} \alpha_r \sigma^r, \quad . \quad . \quad . \quad (1.2)$$

holomorphic at all points within the unit circle C , of which the boundary equation in complex co-ordinates $\sigma, \bar{\sigma}$ is

$$\sigma \bar{\sigma} = 1, \quad . \quad . \quad . \quad (1.3)$$

and satisfying a boundary condition

$$2i\psi = \omega(\sigma) - \bar{\omega}(\bar{\sigma}) = F(\sigma, \bar{\sigma}), \quad . \quad . \quad . \quad (1.4)$$

where $F(\sigma, \bar{\sigma})$ is a given function, a pure imaginary.

Now make use of the boundary equation (1.3) to express the boundary function in terms of σ only, as a series

$$F(\sigma, \bar{\sigma}) = F(\sigma, \sigma^{-1}) = \sum_{-\infty}^{\infty} \beta_r \sigma^r. \quad . \quad . \quad . \quad (1.5)$$

Then, using (1.2), (1.4) becomes

$$\sum_0^{\infty} \alpha_r \sigma^r - \sum_0^{\infty} \bar{\alpha}_r \sigma^{-r} = \sum_{-\infty}^{\infty} \beta_r \sigma^r,$$

† Communicated by the Author.

whence $\alpha_r = \beta_r, \quad \bar{\alpha}_r = -\bar{\beta}_{-r}, \quad \alpha_0 - \bar{\alpha}_0 = \beta_0,$

so that $\omega(\sigma) = \alpha_0 + \sum_1^{\infty} \beta_r \sigma^r, \quad (1.6)$

and so is determined apart from a real constant, which is usually immaterial in the physical applications of the complex potential for simply connected cross-sections. Note that once the boundary function has been obtained in series form, the complex potential can be written down at once as the part of the series with positive powers of σ . The method enables us to write the boundary function $F(\sigma, \bar{\sigma})$ in the separable form $f(\sigma) - f(\bar{\sigma})$, where $f(\sigma)$ is holomorphic, and so pick out the complex potential $\omega(\sigma)$ as $f(\sigma)$ by inspection; in this light the method is an extension of the method employed by the writer previously⁽³⁾ in the z -plane, for those boundaries, in the case of the torsion problem, in which the boundary function $iz\bar{z}$ could be rendered separable in z and \bar{z} by a use of the boundary equation in complex co-ordinates.

The method is equally well adapted to deal with harmonic problems relating to the exterior of a simple boundary capable of being mapped conformally on the exterior of unit circle in the σ -plane by a transformation

$$z = z(\sigma) = \sum_{-1}^{\infty} a_r \sigma^{-r}, \quad a_{-1} \neq 0, \quad (1.7)$$

where $z'(\sigma)$ does not vanish or become infinite at any point within the domain, including the point at infinity, hence the condition $a_{-1} \neq 0$.

The complex potential

$$\omega(\sigma) = \sum_0^{\infty} \alpha_r \sigma^{-r} (1.8)$$

holomorphic outside unit circle, is then given, by a similar treatment, as

$$\omega(\sigma) = \alpha_0 - \sum_0^{\infty} \bar{\beta}_r \sigma^{-r}, \quad (1.9)$$

and is again determined apart from a real constant †.

The results (1.6) and (1.9) might also have been obtained from the boundary condition in the form

$$\omega(\zeta) = \sum_{-\infty}^{\infty} \beta_r \zeta^r + \bar{\omega}(\zeta^{-1})$$

along the unit circle for which $\sigma = \zeta = e^{i\theta}$ by the device, used by Muschelisvili⁽⁵⁾ in biharmonic problems, of multiplying each side by

$$\begin{aligned} \frac{1}{2\pi i} \frac{d\zeta}{\zeta - \sigma} &= \frac{1}{2\pi i} \sum_0^{\infty} \left(\frac{\sigma}{\zeta} \right)^r \frac{d\zeta}{\zeta}, \\ \text{or} \quad &= - \frac{1}{2\pi i} \sum_1^{\infty} \left(\frac{\zeta}{\sigma} \right)^r \frac{d\zeta}{\zeta}, \end{aligned}$$

according as we are dealing with a problem internal or external to the boundary C_1 , so that σ is the affix of a point internal or external to unit

† This case has been considered recently by Lewis⁽⁴⁾.

circle C as the case may be. We then integrate completely round C, and a use of Cauchy's theorem and the residue theorem gives the same results (1.6) and (1.9). The procedure employed earlier is rather more simple, when the series form is readily obtainable.

2. The Torsion Problem.

In the torsion problem for a cylinder of uniform cross-section S_1 , the displacements u, v, w , where w is parallel to the Z -axis and generators of the cylinder, and the non-vanishing stresses \widehat{xz} and \widehat{yz} are given in complex combination as

$$u + iv = i\tau z Z, \quad w = \frac{1}{2}\tau\{\omega_1(z) + \bar{\omega}_1(\bar{z})\}, \quad \dots \quad (2.1)$$

$$\Psi = \widehat{xz} + i\widehat{yz} = \mu\tau\{\bar{\omega}_1'(z) + iz\}, \quad \dots \quad (2.2)$$

where μ is the rigidity and τ is the constant twist per unit length, whilst the complex potential $\omega_1(z)$ satisfies a boundary condition

$$\omega_1(z) - \bar{\omega}_1(\bar{z}) = iz\bar{z} + \text{const.}, \quad \dots \quad (2.3)$$

where the constant can conveniently be taken to be zero when the boundary is a single closed curve as in the present case. The twisting couples N , which must be oppositely applied at the ends of the cylinder, are given as the real part of

$$N + iN^* = i \int_{S_1} z \bar{\Psi} dS_1. \quad \dots \quad (2.4)$$

Now, using (1.1),

$$F(\sigma, \sigma^{-1}) = iz(\sigma)\bar{z}(\sigma^{-1}) = i \sum_0^\infty a_r \sigma^r \sum_0^\infty \bar{a}_s \sigma^{-s} = i \sum_{-\infty}^{+\infty} b_r \sigma^r, \quad \dots \quad (2.5)$$

$$\text{where} \quad b_r = \sum_0 a_{r+s} \bar{a}_s, \quad b_{-r} = \bar{b}_r, \quad r = 0, 1, 2, \dots, \infty, \quad \dots \quad (2.6)$$

and by the method of the last section the solution for the complex potential is immediately obtained as

$$\omega(\sigma) = i \sum_1^\infty b_r \sigma^r. \quad \dots \quad (2.7)$$

Also, from (2.2), (2.4) can be written

$$N + iN^* = \frac{1}{2}i\mu\tau \int_{S_1} \frac{\partial}{\partial \bar{z}} \{2z\bar{z}\omega_1'(z) - iz\bar{z}^2\} dS,$$

or, using the complex form of Stokes's theorem (see Milne-Thomson, 'Theoretical Hydrodynamics,' p. 238 (1938)), N is the real part of

$$N + iN^* = \frac{1}{4}\mu\tau \int_{C_1} z\bar{z}\{2\omega_1'(z) - i\bar{z}\} dz,$$

$$\text{or} \quad N + iN^* = \frac{1}{4}\mu\tau \int_C z(\sigma)\bar{z}(\sigma^{-1})\{2\omega'(\sigma) - i\bar{z}(\sigma^{-1})z'(\sigma)\} d\sigma. \quad \dots \quad (2.8)$$

Now from (1.1)

$$i\bar{z}(\sigma^{-1})z'(\sigma) = i \sum_0^\infty \bar{a}_r \sigma^{-r} \sum_0^\infty s a_s \sigma^{s-1} = i \sum_{-\infty}^\infty A_n \sigma^{n-1}, \quad \dots \quad (2.9)$$

where, if
$$c_n = \sum_{r=1}^{\infty} r a_{n+r} \bar{a}_r, \quad \dots \dots \dots (2.10)$$

$$A_n = n b_n + c_n, \quad A_{-n} = \bar{c}_n, \quad n = 0, 1, 2, \dots \dots \infty \quad \dots \dots (2.11)$$

Hence, using (2.6) and (2.8)–(2.11), we have

$$N + iN^* = \frac{1}{4}\mu\tau \int_C \left\{ \sum_{-\infty}^{+\infty} b_s \sigma^s \right\} \sum_1^{\infty} \{ (r b_r - c_r) \sigma^r - \bar{c}_r \sigma^{-r} \} \frac{d\sigma}{\sigma},$$

and, using the residue theorem, this becomes

$$N = \frac{1}{2}\pi\mu\tau \left\{ b_0 c_0 + \sum_1^{\infty} (b_r \bar{c}_r + \bar{b}_r c_r - r b_r \bar{b}_r) \right\}, \quad \dots \dots (2.12)$$

which is effectively equivalent to Morris's solution of the problem ⁽²⁾.

3. The Fluted Column.

As a particular example of this somewhat general form of solution, consider the boundary given by $\xi=0$ in the curvilinear system of co-ordinates ξ, η given by

$$\zeta = \xi + i\eta, \quad \sigma = e^{\zeta}, \quad z = c\sigma(1 + \lambda\sigma^n), \quad \dots \dots (3.1)$$

shown elsewhere ⁽³⁾ to be a regular curvilinear polygon of n "sides," so that the cylinder is a grooved or fluted column of a special type.

With our previous notation of (1.1) the non-vanishing coefficients a_r are

$$a_1 = c, \quad a_{n+1} = \lambda c.$$

Again from (2.6)
$$b_r = \bar{a}_1 a_{r+1} + \bar{a}_{n+1} a_{n+r+1},$$

so that the non-vanishing coefficients b_r are

$$b_0 = a_1 \bar{a}_1 + a_{n+1} \bar{a}_{n+1} = c^2(1 + \lambda^2), \quad b_n = \bar{a}_1 a_{n+1} = \lambda c^2.$$

It follows that the complex potential for the problem is

$$\omega(\sigma) = i\lambda c^2 \sigma^n. \quad \dots \dots \dots (3.2)$$

Further, (2.10) gives
$$c_r = \bar{a}_1 a_{r+1} + (n+1) \bar{a}_{n+1} a_{n+r+1},$$

whence the non-vanishing coefficients c_r are

$$c_0 = \bar{a}_1 a_1 + (n+1) \bar{a}_{n+1} a_{n+1} = c^2 \{ 1 + (n+1) \lambda^2 \}, \quad c_n = \bar{a}_1 a_{n+1} = \lambda c^2.$$

Finally, from (2.12), the torsion moment for the fluted column is given by

$$N = \frac{1}{2}\pi\mu\tau c^4 \{ 1 + 4\lambda^2 + (n+1)\lambda^4 \}. \quad \dots \dots \dots (3.3)$$

In particular ⁽³⁾ the case $n=1$ gives the boundary as the elliptic limaçon

$$r = a + b \cos \theta, \quad b \leq a, \quad (a=c, \quad b/a=2\lambda).$$

In this case we have from (3.3)

$$N = \frac{1}{16}\pi\mu\tau \{ 8a^4 + 8a^2b^2 + b^4 \}, \quad \dots \dots \dots (3.4)$$

agreeing with the value previously found ⁽⁶⁾, and putting $b=a$, we have the known result for the cardioid cross-section ⁽⁷⁾.

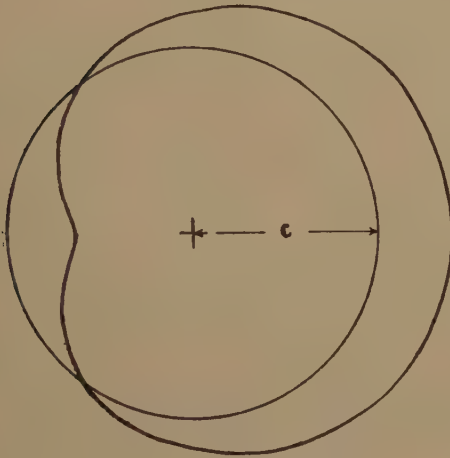
The peripheral shear $\widehat{\eta z}$ is readily obtained, since $\widehat{\xi z}$ vanishes on the boundary $\xi=0$, from the complex combination $\Psi'=\widehat{\xi z}+i\widehat{\eta z}$, as

$$\Psi'=e^{-i\alpha}\Psi, \quad \text{where } e^{2i\alpha}=\frac{z'(\zeta)}{\bar{z}'(\zeta)}=\frac{\sigma z'(\sigma)}{\bar{\sigma}\bar{z}'(\bar{\sigma})},$$

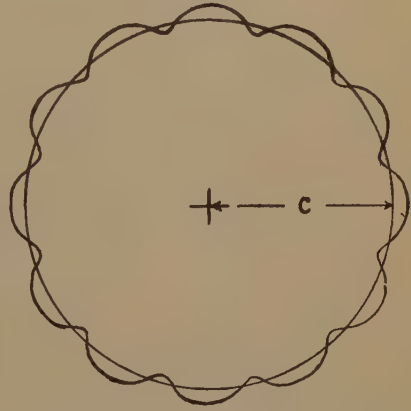
and
$$\bar{\Psi}=\mu\tau\left\{\frac{\omega'(\sigma)}{z'(\sigma)}-i\bar{z}(\sigma^{-1})\right\}, \quad \dots \dots \dots (3.5)$$

whence
$$\widehat{\eta z}=\frac{\mu\tau\sigma}{\{z'(\sigma)\bar{z}'(\sigma^{-1})\}^{1/2}}\{z'(\sigma)\bar{z}(\sigma^{-1})+i\omega'(\sigma)\},$$

which leads to
$$\widehat{\eta z}=\mu\tau c \frac{1+\lambda^2(n+1)+2\lambda\cos n\eta}{\{1+\lambda^2(n+1)^2+2\lambda(n+1)\cos n\eta\}^{1/2}}. \quad \dots \dots \dots (3.6)$$



(i.) $n=1, \quad \lambda=3/8(<1/2)$.



(ii.) $n=12, \quad \lambda=1/16(<1/13)$.

Examples of the boundary curve C_0 .

Notice, from (3.5), that the stresses are necessarily finite across the cross-section, since from (1.1), $z'(\sigma)\neq 0$ at any point within the cross-section.

At the bottom of a groove, where r is a minimum, $\cos n\eta=-1$ and we have

$$\widehat{\eta z}=\mu\tau c \frac{(1-\lambda)^2+n\lambda^2}{1-\lambda(n+1)}, \quad \dots \dots \dots (3.7)$$

becoming infinite when the groove becomes a cusp, i. e., when $\lambda(n+1)=1$. At the points where r becomes a maximum, $\cos n\eta=1$ and we have

$$\widehat{\eta z}=\mu\tau c \frac{(1+\lambda)^2+n\lambda^2}{1+\lambda(n+1)}. \quad \dots \dots \dots (3.8)$$

For the dodecagonal form of the boundary described in the previous paper⁽³⁾, where $n=12, \lambda=1/16$, these become

$$\text{for } r_{\min} \widehat{\eta z}=\frac{79}{16}\mu\tau c \simeq 5\mu\tau c, \quad \text{for } r_{\max} \widehat{\eta z}=\frac{301}{464}\mu\tau c \simeq \frac{2}{3}\mu\tau c,$$

instructive simple numerical results showing the considerable variation in peripheral shear due to the fluting of the column; for the circular cylinder of radius c the shear would have the constant value μrc . If N_0 is the torsion moment for this circular cylinder, obtained from (3.3) by putting $\lambda=0$, the torsion moment for the dodecagonal fluted column is

$$N=1.01578 N_0.$$

References.

- (1) Stevenson, Phil. Trans. Roy. Soc. cccxxvii. p. 161 (1938).
- (2) Morris, *Math. Annalen*, cxvi. p. 374 (1939).
- (3) Stevenson, Phil. Mag. xxxiv. p. 105 (1943).
- (4) Lewis, Quart. J. of Math. (Oxford) xii. p. 30 (1941).
- (5) Muschelishvili, *Z. A. M. M.* xiii. p. 264 (1933).
- (6) Stevenson, Proc. Lond. Math. Soc. xlv. p. 126 (1939).
- (7) Shepherd, 'Engineering,' cxxviii. p. 39 (1929).

XVI. The Repulsive Action between a Heated Solid Surface and the Adjacent Gas.—Part I.

By H. SPENCER-GREGORY, Ph.D., A.R.C.S., D.I.C.,
Imperial College of Science and Technology *.

[Received October 9, 1942.]

It may be generally stated that in all cases in which heated solid surfaces are disposed in an enclosure at a lower temperature thermo-molecular forces are set up between any element of the surface and the adjacent gas.

An important example of such action was considered by Knudsen ⁽¹⁾ in relation to the repulsion exerted between two uniformly heated parallel surfaces disposed in a strongly rarefied gas and theoretically represented by the expression,

$$F=\frac{1}{2}.p.\left(\sqrt{\frac{T_1}{T_2}}-1\right), \quad \dots \quad (1)$$

where p denotes the gas pressure, T_1 and T_2 the absolute temperatures of the opposing surfaces, and F the radiometric pressure per unit area of surface.

Equation (1) was modified by von Smoluchowski ⁽²⁾ in order to include the effect of thermal accommodation and now takes the form

$$F=\frac{1}{2}.p.\left[\left(\sqrt{\frac{T_1}{T_2}}-1\right)+\frac{(1-\alpha)}{4(2-\alpha)^2}\left(\frac{T_1}{T_2}-1\right)^2\right], \quad \dots \quad (2)$$

and in which α , the coefficient of thermal accommodation as defined by Knudsen, is assumed to be the same for both surfaces.

* Communicated by the Author.

Apart from the difficulties associated with the preparation of identical surface conditions, equation (2) is incorrect from reasons of the dependence of α on temperature.

For gases in a state of partial rarefaction equations (1) and (2) are no longer valid, and the repulsive action for the same parallel plate system is now one strictly between either surface and the gas adjacent to it. For uniformly heated parallel surfaces, the repulsive force F per unit area is expressible in the form

$$F = \frac{1}{2}P \left[\sqrt{1 + \frac{2 \cdot \alpha \cdot \delta\theta}{(2-\alpha)\left(T_1 - \frac{2 \cdot \delta\theta}{(2-\alpha)}\right)}} - 1 \right] \dots \dots (3)$$

in which T_1 denotes the temperature of the uniformly heated surface and $\delta\theta$ the temperature discontinuity between the surface and the adjacent gas. For a parallel plate distribution of temperature, $\delta\theta$ takes the form ⁽³⁾

$$\begin{aligned} \delta\theta &= f\left(\frac{Q}{P}\right) \\ &= \frac{Q\sqrt{T_1}}{pA} \left\{ \sqrt{1 + \frac{Q^2 \cdot T_1}{p^2 A^2 T_1^2 (2-\alpha)^2}} - \frac{Q^2 \sqrt{T_1}}{p \cdot A (2-\alpha) T_1} \right\} \dots \dots (4) \end{aligned}$$

where Q represents the rate of heat transfer per unit area through the gas, p the gas pressure, and A a constant defined by

$$A = (\beta + \frac{1}{2}) \frac{2\alpha}{(2-\alpha)} \sqrt{\frac{R}{2 \cdot \pi \cdot M}} \dots \dots \dots (5)$$

in which $\beta \cdot R$ denotes the molecular heat of the gas and M its molecular weight.

If inequalities of temperature exist over the surface then, as contended by Maxwell ⁽⁴⁾, the major portion of the surface is radiometrically inactive, and the repulsive action is then found to be due to mass motions of the gas set up by the temperature differences existing on the surface.

Such a streaming condition of the gas according to Maxwell and to Hettner ⁽⁵⁾ is theoretically represented by the equation

$$u - \gamma_M du/dx_1 = \frac{3\eta}{4 \cdot \rho \cdot T} dT/dx_2 \dots \dots \dots (6)$$

in which u is the velocity of tangential flow at a point on the surface where the temperature gradient is represented by dT/dx_2 , γ_M denotes the coefficient of slip, ρ the gas density, and x_1 the normal direction to the surface.

An experimentally modified form of equation (6) was used by Weber ⁽⁶⁾ in order to calculate the radiometer pressure associated with parallel plate radiometers of the guard ring type.

This equation may be expressed in the form,

$$u - k_2 du/dx_1 = \frac{3}{4} \cdot k_1 \frac{\eta}{\rho \cdot T} dT/dx_2 \frac{1}{1 + \lambda/d} \dots \dots \dots (7)$$

where $k_2 = \frac{2-s}{s}$.

s being the coefficient of molecular momentum transfer and k_1 a constant of approximate value $\frac{2}{3}$. The factor $1 + \lambda/d$ is of empirical value in Weber's equation, but may be justified theoretically from considerations of temperature drop theory, and in the author's opinion is more correctly expressed in the form $1 + q \cdot \lambda/T \cdot d$, where d is the distance between the plate system, λ the mean free path, and q a constant. Calculations of the radiometer pressure were carried out by Weber in terms of the thermo-molecular pressure gradient deduced from (7) in conjunction with the appropriate equation expressing the velocity distribution of the gas enclosed by the parallel plate system.

Weber's formula⁽⁷⁾ for the radiometric pressure F per unit area may be expressed in the form,

$$F = \frac{1}{2 \cdot T} \cdot \frac{1}{\alpha_1 \left(\frac{d}{\lambda_1}\right)^2 p + \beta_1 \left(\frac{d}{\lambda_1}\right) + \mu \frac{1}{p}} \quad \dots \quad (8)$$

where T denotes the average temperature between the plates, p the gas pressure, and α_1 , β_1 , and μ are constants.

Weber's theory has been tested recently by Fredlund⁽⁸⁾ in an elaborate series of experiments and found to represent the experimental data with fair accuracy except in the range where the gases are strongly rarefied and also at the higher limits of pressure beyond which the effect becomes negligibly small. Fredlund⁽⁸⁾ points out a serious discrepancy in the case of hydrogen, a gas for which the influence of thermal accommodation would be of serious consequence in relation to (8).

In the present instance the author considers an extension of the temperature drop theory in order to calculate the repulsive action between a uniformly heated thin wire and the adjacent gas, and it will be shown that such forces of repulsion are expressible in terms of the quantity Q/p , where Q represents the rate of heat loss from the wire and p the pressure of the gas.

For a state of thermal equilibrium (fig. 1) the molecules arriving at the wire surface possess an average temperature T_1'' , and the number of these crossing the area $2 \cdot \pi(r_1 + \lambda_1) \cdot L$ is represented by the expression

$$2 \cdot \pi(r_1 + \lambda_1) \cdot L \cdot \frac{n \cdot C_1''}{\sqrt{6 \cdot \pi}},$$

where L is the length of the heated filament of temperature T_1 and n and C_1'' the molecular concentration and root mean square velocity of the gas molecules at a temperature T_1'' and λ_1 the mean free path.

The number colliding or rebounding from the filament per sec. is therefore

$$\begin{aligned} & \frac{2\pi(r_1 + \lambda_1)L \cdot n \cdot C_1''}{\sqrt{6 \cdot \pi}} \cdot \frac{r_1}{r_1 + \lambda_1}, \\ & = 2 \cdot \pi r_1 L \cdot \frac{n \cdot C_1''}{\sqrt{6 \cdot \pi}}. \end{aligned}$$

The average normal momentum possessed by each incident molecule is represented, in terms of the kinetic theory, by the integral

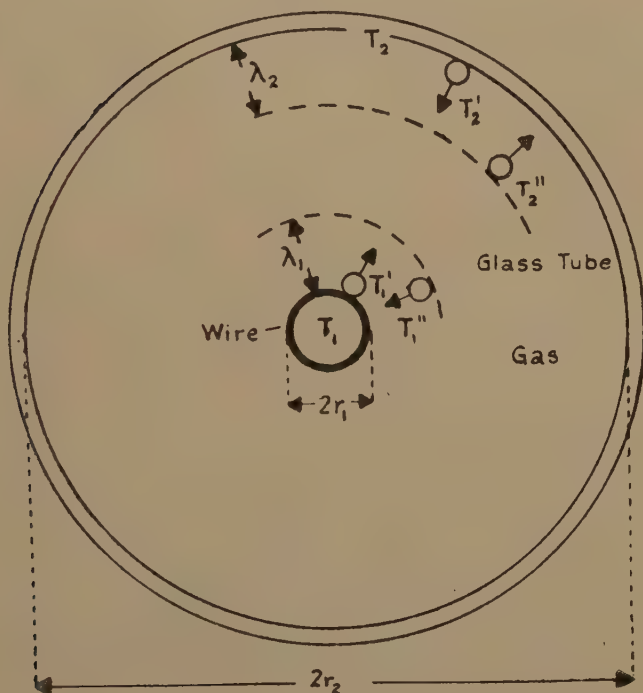
$$\frac{1}{n \cdot C_1'' / \sqrt{6\pi}} \int_0^{\pi/2} \frac{\sin \theta \cos^2 \theta d\theta}{2} \int_0^\infty dn_c \cdot mc^2$$

$$= mC_1'' \sqrt{\frac{\pi}{6}},$$

m being the mass of a molecule and

$$dn_c \equiv 4n \sqrt{\frac{h^3 m^3}{\pi}} e^{-hmc^2/c^2} dc;$$

Fig. 1.



hence the total pressure exerted by the incident molecules on the surface of the wire is given by

$$2\pi \cdot r_1 \cdot L \cdot \frac{n \cdot C_1''}{\sqrt{6\pi}} \cdot m \cdot C_1'' \sqrt{\frac{\pi}{6}}.$$

A similar expression holds for an equal number emitted from the surface, and so we obtain for the total pressure per unit area of the surface

$$p' = \frac{nC_1''}{\sqrt{6\pi}} \left[m \cdot C_1'' \sqrt{\frac{\pi}{6}} + m \cdot C_1' \sqrt{\frac{\pi}{6}} \right],$$

$$\begin{aligned}
 &= \frac{m \cdot n C_1''^2}{6} \left[1 + \frac{C_1'}{C_1''} \right] \\
 &= \frac{1}{2} \cdot p \cdot \left(1 + \frac{C_1'}{C_1''} \right), \quad \dots \dots \dots (9)
 \end{aligned}$$

where p denotes the pressure of the gas.

The repulsive force between the wire and the adjacent gas per unit area of the surface of the wire is therefore

$$\begin{aligned}
 F' &= p' - p = \frac{1}{2} p \cdot \left(\frac{C_1'}{C_1''} - 1 \right), \\
 &= \frac{1}{2} \cdot p \cdot \left(\sqrt{\frac{T_1'}{T_1''}} - 1 \right), \quad \dots \dots \dots (10)
 \end{aligned}$$

where C_1' and C_1'' represent the velocities of agitation of the molecules rebounding and impinging on the surface of the wire, and whose corresponding mean temperatures are T_1' and T_1'' .

If α represents the coefficient of thermal accommodation as defined by Knudsen, then

$$\alpha = \frac{T_1' - T_1''}{T_1 - T_1''},$$

and since also, ⁽⁹⁾ $T_1 - T_1'' = \frac{2 \cdot \delta\theta}{(2 - \alpha)} \dots \dots \dots (11)$

hence if F represents the total repulsive force between the wire and the adjacent gas we obtain from (9), (10) and (11)

$$F = \pi \cdot r_1 \cdot L \cdot p \cdot \left[\sqrt{1 + \frac{2\alpha\delta\theta}{(2-\alpha)\left(T_1 - \frac{2\delta\theta}{(2-\alpha)}\right)}} - 1 \right] \dots \dots (12)$$

Solving (12) for $\delta\theta$ we have, substituting F' for $F/2\pi r_1 L$, the expression

$$\begin{aligned}
 \delta\theta &= \frac{T_1 \left(1 - \frac{\alpha}{2} \right) \left(\frac{2F'}{p} \right) \left(\frac{2F'}{p} + 2 \right)}{\alpha + \left(\frac{2F'}{p} \right) \left(\frac{2F'}{p} + 2 \right)} \dots \dots \dots (13)
 \end{aligned}$$

For a pressure range over which the curvature of the wire is large compared with the free path, equation (12) reduces to the form

$$\begin{aligned}
 \delta\theta &= \frac{F \cdot T_1 (2 - \alpha)}{p \cdot \pi \cdot r_1 \cdot L \cdot \alpha}, \\
 &= \frac{F \cdot T_1}{p \cdot r_1 \cdot B} \dots \dots \dots (14)
 \end{aligned}$$

where B denotes the factor $\frac{\pi L \alpha}{(2 - \alpha)}$.

The temperature discontinuity between the wire and gas is expressible therefore in terms of the quantity F/p .

Since also $\delta\theta$ may be expressed in the form

$$\delta\theta = \frac{Q\sqrt{T_1}}{p \cdot r_1 \cdot A}, \quad \dots \quad (15)$$

where Q represents the rate of heat transport through the gas, we obtain from (14) and (15)

$$\begin{aligned} Q/F &= A/B = (\beta + \frac{1}{2}) \sqrt{\frac{8RT_1}{\pi M}}, \\ &= (\beta + \frac{1}{2}) \bar{v} = \gamma, \end{aligned}$$

where \bar{v} denotes the arithmetic mean velocity of agitation and γ a constant depending only on temperature.

For a cylindrical distribution of temperature (fig. 1), Fourier's law for the rate of heat transport from the wire, modified in respect of temperature drop, is represented by the equation

$$Q \log_c \frac{r_2}{r_1} = 2 \cdot \pi \cdot \kappa \cdot L \theta - 2 \cdot \pi \cdot \kappa \cdot L (\delta\theta_1 + \delta\theta_2), \quad \dots \quad (17)$$

where $\delta\theta_1$ and $\delta\theta_2$ denote the temperature discontinuities at the surfaces of the wire and tube, r_1 and r_2 the radii of the wire and internal surface of the tube, k the thermal conductivity of the gas at a temperature $(T_1 + T_2)/2$, L the length of the uniformly heated wire, and θ the temperature difference between the wire and the internal surface of the tube.

Since also

$$Q = \gamma_1 F_1 = \gamma_2 F_2,$$

where F_1 and F_2 represent the forces of repulsion at the surfaces of the wire and tube wall, and γ_1 and γ_2 the corresponding coefficients.

Hence, by substituting

$$Q = \gamma_1 F_1$$

and

$$\delta\theta_1 = \frac{F_1 \cdot T_1}{p \cdot r_1 B_1}, \quad \delta\theta_2 = \frac{F_2 T_2}{p r_2 B_2} \quad \text{in (17),}$$

we obtain

$$\gamma_1 F_1 \log_c \frac{r_2}{r_1} = 2\pi k L \theta - 2\pi k L \left(\frac{F_1 T_1}{p r_1 B_1} + \frac{F_2 T_2}{p \cdot r_2 B_2} \right) \quad \dots \quad (18)$$

Since

$$F_2 = F_1 \frac{\gamma_1}{\gamma_2} = F_1 \sqrt{\frac{T_1}{T_2}},$$

and for moderate values of the difference of temperature $T_1 - T_2$ we may also with sufficient accuracy write $B_2 = B_1$, and finally (18) takes the form

$$\frac{1}{F_1} = \frac{\log_c \frac{r_2}{r_1}}{2\pi \cdot f \cdot L \cdot \theta} + \frac{\sqrt{T_1}}{p \cdot \theta \cdot B_1} \left(\frac{\sqrt{T_1}}{r_1} + \frac{\sqrt{T_2}}{r_2} \right), \quad \dots \quad (19)$$

where f is written for the term k/γ_1 .

The quantity f (dynes. cm.⁻¹ deg.⁻¹) termed by the author "the coefficient of repulsive action" is therefore independent of pressure and depends only on the temperature through k and γ_1 .

The quantities $1/Q$ and $1/F$, satisfy a linear relation with $1/p$, and simultaneous observations of the quantities $1/Q$ and $1/F$, and their variation with the gas pressure are sufficient to determine the values of k , α , and $\beta.R$.

Corresponding theoretically derived expressions in relation to the repulsive action between a uniformly heated sphere and the adjacent gas when surrounded by a concentrically disposed outer spherical surface are as follows:—

Temperature discontinuity $\delta\theta_s$,

$$\delta\theta_s = \frac{Q\sqrt{T_1}}{p \cdot r_1 A_s} \left[\sqrt{1 + \frac{Q \cdot T_1}{(2-\alpha)^2 p^2 r_1^4 A_s^2 T_1^2}} - \frac{Q \cdot \sqrt{T_1}}{(2-\alpha) \cdot p \cdot r_1^2 A_s T_1} \right],$$

where r_1 and r_2 are the radii of the inner and outer spherical surfaces and T_1 and T_2 their temperatures (abs.).

A_s is a constant defined by the relation

$$A_s = 4\pi \sqrt{\frac{R}{2\pi M}} (\beta + \frac{1}{2})^{\frac{2\alpha}{2-\alpha}}, \quad \dots \dots \dots (21)$$

and Q the rate of heat transfer through the gas.

An approximate value for $\delta\theta_s$ is given by

$$\delta\theta_s = \frac{Q\sqrt{T_1}}{p \cdot r_1^2 \cdot A_s} \dots \dots \dots (22)$$

If F denotes the repulsive force between the inner sphere and the adjacent gas, then it can be shown that

$$F = 4\pi r_1^2 \frac{p}{2} \left[\sqrt{1 + \frac{2 \cdot \alpha \cdot \delta\theta_s}{(2-\alpha) \left(T_1 - \frac{2\delta\theta_s}{2-\alpha} \right)}} - 1 \right] \dots \dots \dots (23)$$

hence, from (23) we obtain for $\delta\theta_s$,

$$\begin{aligned} \delta\theta_s &= \frac{\left(\frac{2F'}{p} + 2 \right) \left(\frac{2F'}{p} \right) \frac{(2-\alpha)T_1}{\alpha}}{\alpha + \left(\frac{2F'}{p} + 2 \right) \left(\frac{2F'}{p} \right)} \\ &= \frac{F \cdot T_1}{p r_1^2 B_s} \quad \text{approximately,} \quad \dots \dots \dots (24) \end{aligned}$$

where $F' = \frac{F}{4\pi r_1^2}$,

and B_s is written for the expression $\frac{2\pi \cdot \alpha}{2-\alpha}$.

Since also approximately from (20)

$$\delta\theta_s = \frac{Q\sqrt{T_1}}{p \cdot r_1^2 \cdot A_s},$$

we obtain therefore

$$Q/F = (\beta + \frac{1}{2}) \sqrt{\frac{8RT_1}{\pi M}} = \gamma \quad \dots \quad (25)$$

For a spherical distribution of temperature, Fourier's law of heat conduction, modified in relation to the temperature drop effect takes the form

$$\frac{\theta}{Q} = \frac{(r_2 - r_1)}{4\pi \cdot K \cdot r_1 \cdot r_2} + \frac{1}{pA_s} \left(\frac{\sqrt{T_1}}{r_1^2} + \frac{\sqrt{T_2}}{r_2^2} \right), \quad \dots \quad (26)$$

which in conjunction with (25) becomes

$$\frac{\theta}{F} = \frac{(r_2 - r_1)}{4\pi \cdot f \cdot r_1 \cdot r_2} + \frac{\sqrt{T_1(2-\alpha)}}{2\pi \cdot \alpha \cdot p} \left(\frac{\sqrt{T_1}}{r_1^2} + \frac{\sqrt{T_2}}{r_2^2} \right). \quad (27)$$

Similarly, expressions may be derived for the case of a parallel plate distribution of temperature in which the surface temperatures are T_1 and T_2 respectively

The temperature discontinuity between the hot plate T_1 and the gas is represented by the relation

$$\begin{aligned} \delta\theta &= \frac{Q \cdot \sqrt{T_1}}{p \cdot A} \left\{ \sqrt{1 + \frac{Q^2 \cdot T_1^2}{p^2 \cdot A^2 T_1^2 (2-\alpha)^2}} - \frac{Q \sqrt{T_1}}{p \cdot A (2-\alpha) T_1} \right\} \quad \dots \quad (28) \\ &= \frac{Q \cdot \sqrt{T_1}}{p \cdot A} \quad \text{approximately.} \end{aligned}$$

In (28) Q denotes the rate of heat transport through the gas per unit area and A is defined by

$$A = (\beta + \frac{1}{2}) \frac{2\alpha}{(2-\alpha)} \sqrt{\frac{R}{2 \cdot \pi \cdot M}}.$$

The repulsive force per unit area between T_1 and the adjacent gas is then represented by the expression

$$F = \frac{1}{2} \cdot p \cdot \left[\sqrt{1 + \frac{2 \cdot \alpha \cdot \delta\theta}{(2-\alpha) \left(T_1 - \frac{2 \cdot \delta\theta}{(2-\alpha)} \right)}} - 1 \right] \quad \dots \quad (29)$$

Solving (29) for $\delta\theta$ we obtain

$$\begin{aligned} \delta\theta &= \frac{T_1(2-\alpha) \left(\frac{F}{p} + 1 \right) \frac{2F}{p}}{\alpha + \left(\frac{2F}{p} + 2 \right) \frac{2F}{p}}, \\ &= \frac{(2-\alpha)T_1}{\alpha} \frac{2F}{p} \quad \text{approximately,} \quad \dots \quad (30) \end{aligned}$$

hence from (28) and (30) we have

$$\frac{Q}{F} = \sqrt{\frac{T_1 \cdot 8 \cdot R}{\pi \cdot M}} (\beta + \frac{1}{2}) = v(\beta + \frac{1}{2}) = \gamma,$$

as before.

The Fourier equation for the heat transfer modified in respect of the temperature drop effect can be shown to be of the form

$$\frac{(T_1 - T_2)}{Q} = \frac{d}{K} + \frac{1}{p \cdot A} (\sqrt{T_1} + \sqrt{T_2}), \quad \dots \quad (31)$$

where d denotes the distance of separation of the surface and A is assumed to have the same value for both surfaces.

Substituting $Q = \gamma \cdot F$ in (31), it follows that

$$\frac{T_1 - T_2}{F} = \frac{d}{\gamma} + \frac{\gamma}{p \cdot A} (\sqrt{T_1} + \sqrt{T_2}), \quad \dots \quad (32)$$

where f is written for the ratio k/γ and

$$= \frac{\gamma}{A} = \frac{2(2-\alpha)}{\alpha} \sqrt{T_1}.$$

The author has made use of equation (12) in order to calculate the repulsive force F for a range of pressures throughout which the tempera-

Table I. H_2 .

p (cm.) Hg.	F , dynes/cm. ²
·950	79·14
·870	77·41
·728	73·16
·618	68·90
·555	66·11
·463	61·05
·384	56·00
·336	51·90
·272	46·95

Table II. He.

p (cm.) Hg.	F , dynes/cm. ²
·790	157·6
·650	144·6
·560	134·6
·505	128·8
·420	115·7
·350	103·8
·290	92·7
·208	72·2
·180	67·2
·155	61·0
·125	46·3
·115	40·3

ture drop equation (17) is valid. For this purpose the experimental data obtained by Gregory and Stephens⁽¹⁰⁾ relevant to the heat conduction through He and H_2 was used to calculate $\delta\theta$ from (15) in terms of Q the rate of heat transport, from which the corresponding values of F may be determined over the appropriate pressure range. The results of such procedure, tabulated in Tables I. and II., are seen to be comparable in magnitude with the results obtained by Fredlund using a plate radiometer of the Weber type.

Theoretical considerations similar to those which have been applied to the case of the Weber guard ring radiometer are also applicable to the case of an electrically wire disposed coaxially in a tube. The existence of a temperature gradient along the surface of the wire will in general be accompanied by streaming movements of the adjacent gas in a direction from cold to hot, and the effect has been found to depend on the magnitude of the temperature gradient, the dimensions of the wire and tube, and also on the nature and pressure of the gas. The direction of the streaming gas is shown diagrammatically in fig. 2, and

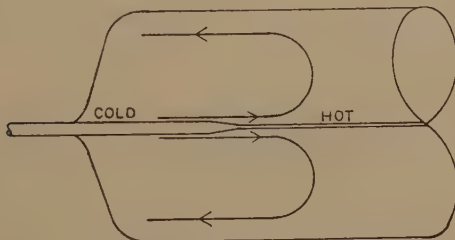
the resulting thermo-molecular pressure gradient set up may be calculated from the Maxwell-Hettner equation, modified in order to include the effects (enhanced in this case) due to slip and temperature drop.

At a point x distant from the junction of the fine and lead wires, the velocity of the gas stream u_x is represented by the relation,

$$u_x = \gamma_M \frac{\partial u}{\partial r_1} + \frac{1/2 \cdot \eta dT/dx}{\rho \cdot T \cdot f\left(\frac{\lambda}{r_1}\right)},$$

where γ_M denotes the coefficient of slip, $\frac{\partial u}{\partial r_1}$ the velocity gradient at a point on the surface of the wire, dT/dx the temperature gradient along the wire at x , $f\left(\frac{\lambda}{r_1}\right)$ a function of the quantity λ/r_1 in which λ and r_1 denote the mean free path and the radius of the wire, and ρ and η are the density and coefficient of viscosity of the gas respectively. If the

Fig. 2.



resulting motion of the gas is viscous in character, the equation expressing the velocity distribution is represented by

$$u = \frac{r^2 dp}{4 \cdot \eta dx} + C \cdot \log r + C',$$

where r is measured in a normal direction from the axis of the wire, and C and C' are constants of integration determined from considerations of the boundary conditions.

Since also the net transfer of gas across any normal cross-section of the tube is zero, we have

$$\int_{r_1}^a 2 \cdot \pi \cdot r \cdot dr \cdot u = 0.$$

where a denotes the radius of the tube.

The problem has also important consequences in relation to the hot wire method of investigating the heat transfer through a gas, and, in particular, as explanitive of the $(Q.p)$ isothermals obtained by Gregory and Archer⁽¹¹⁾ in their experiments on the heat conduction through gases. A solution of the problem as outlined above will be submitted in detail in a subsequent paper.

SUMMARY.

In this part of the paper the problem of the repulsive action between a heated solid and the adjacent gas is discussed generally from aspects when the gas is both partially and strongly rarefied.

The paper is concerned mainly with the case of a uniformly heated thin wire under conditions of partial rarefaction. The repulsive action is shown to depend in magnitude on the temperature discontinuity between the gas and the wire surface. It follows that the reciprocal of the repulsive force $1/F$ satisfies a linear relation with the reciprocal of the pressure $1/p$, a relation which is found to be of the same form as the law of heat conduction through the gas modified in respect of the temperature drop effect.

The results are also given both for a spherical and a parallel plate distribution of temperature, and calculated values of the repulsive forces are tabulated for the gases He and H_2 .

The paper concludes with an outline of the theoretical difficulties arising from the existence of a temperature gradient along the wire surface, together with a summary of the theoretical procedure to be observed, in order to calculate the corresponding forces of repulsion.

References.

- (1) Knudsen, M., *Ann. der Physik*, vol. vi. p. 129 (1930).
- (2) Von Smoluchowski, *Ann. der Physik*, vol. xxxv. p. 1000 (1911).
- (3), (9) H. Spencer-Gregory, *Phil. Mag.* ser. 7, vol. xxii. (1936).
- (4) Maxwell, J. C., *Phil. Trans. Roy. Soc.* vol. clxx. p. 332 (1879).
- (5) Hettner and Czerny, *Zeit. Physik.* vol. xxvii. p. 12 (1924).
- (6), (7) Weber, *Det. Kgl. Danske Videnskabernes Selskab Mathematiskfysiske Meddelelser*, vol. xiv. p. 13 (1937).
- (8) Fredlund, *Phil. Mag.* ser. 7, vol. xxvi. (Dec. 1938).
- (10) Gregory and Stephens, 'Nature,' vol. cxxxix. (Jan. 1937).
- (11) Gregory and Archer, *Proc. Roy. Soc. A*, vol. cx. (1926).

XVII. *Further Notes on the Electron Density Distribution of the Upper Ionosphere.*

By OLOF E. H. RYDBECK, Sc.D.,
Chalmers Institute of Technology, Gothenburg, Sweden*.

[Received April 20, 1942.]

THE problem of obtaining the true electron density distribution from the ionospheric virtual height data has been treated by the author in several communications⁽¹⁾. Since then the true electron density distribution has been computed for a number of ionosphere records. The nature of the electron density distribution of the F_2 -layer has attracted

* Communicated by the Author.

much interest during recent years⁽²⁾. A publication of some of our later results, therefore, should complete the earlier communications.

A number of quiet-day recordings from the Huancayo Observatory at the magnetic equator have kindly been placed at our disposal by the

Fig. 1.

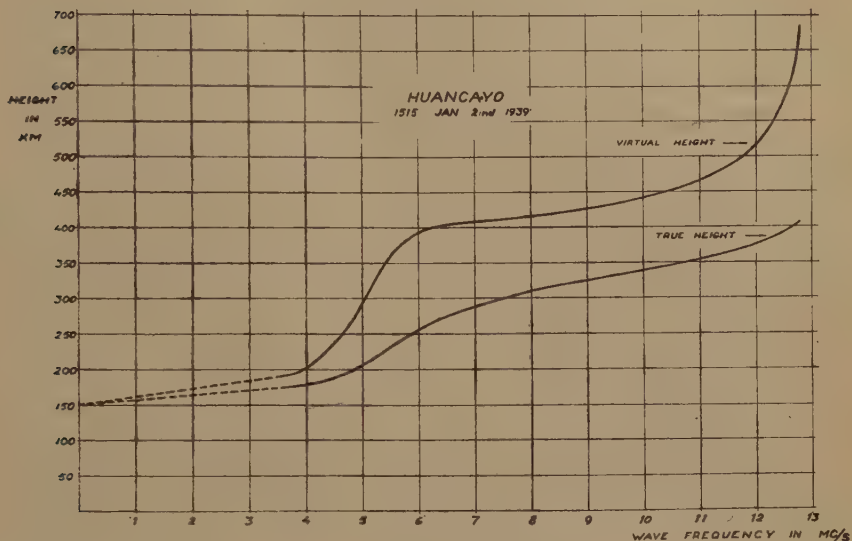
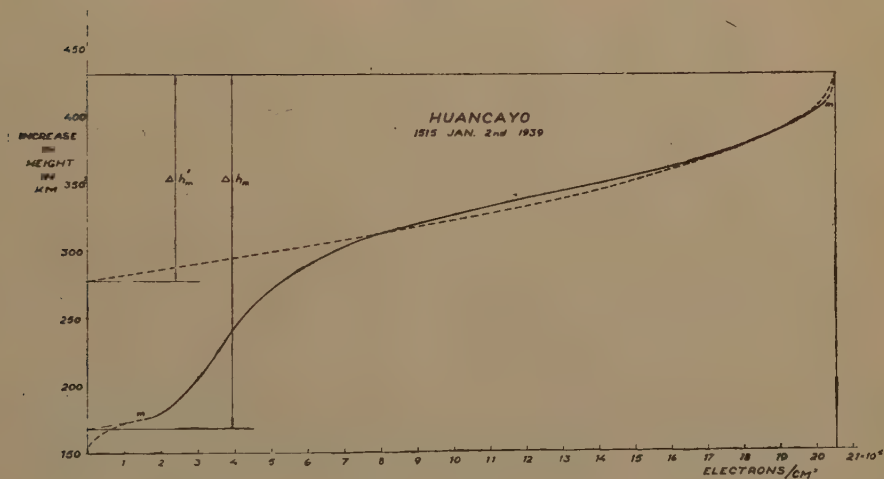


Fig. 2.



director of the Department of Terrestrial Magnetism of the Carnegie Institution, Dr. J. A. Fleming. For several of these the F-layer electron density distribution has been computed from the virtual height data of the ordinary ray.

Fig. 1 shows the virtual and true heights on January 2nd at 15.15 local time. From this the electron density distribution is plotted as shown in fig. 2.

The limits of the reliable region are marked by m . For comparison a dashed parabola has been drawn arbitrarily to fit the upper distribution. This seems to be approximately parabolic over an appreciable range.

Fig. 3.

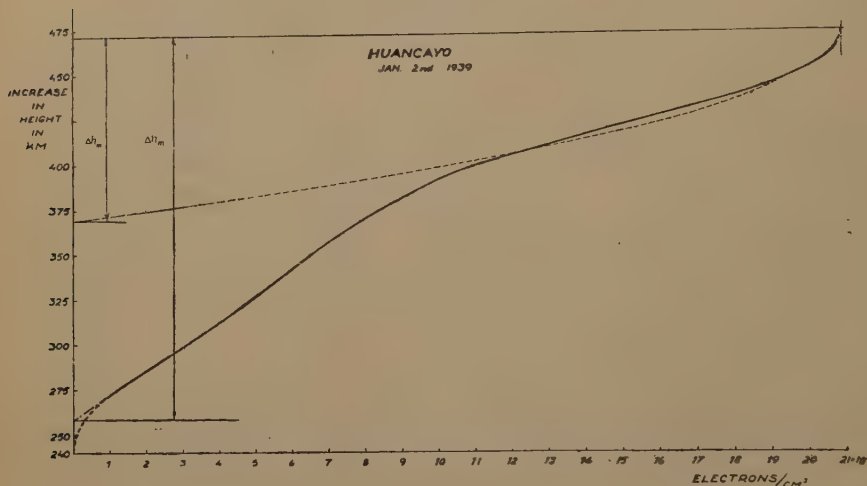
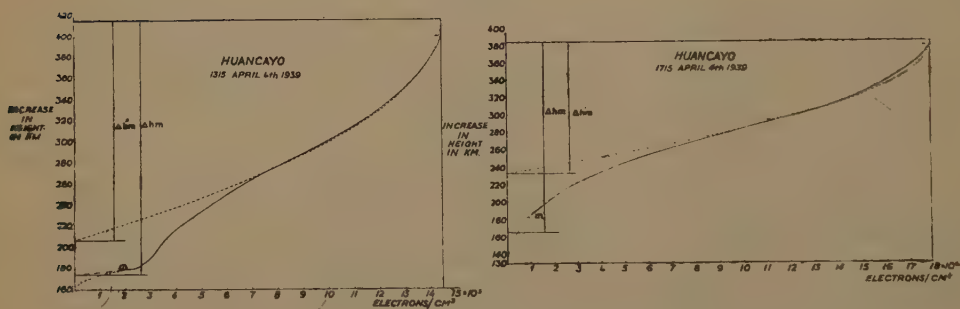


Fig. 4.



For a Chapman layer the electron density in the vicinity of the level of maximum ionization varies as

$$N = N_{\max} \cdot \left(1 - \frac{z^2}{(2H)^2} \right).$$

If a comparison is permissible, $\Delta h'_m/2$ should equal H . This is 76 km. in the present case and should indicate the order of magnitude quite well.

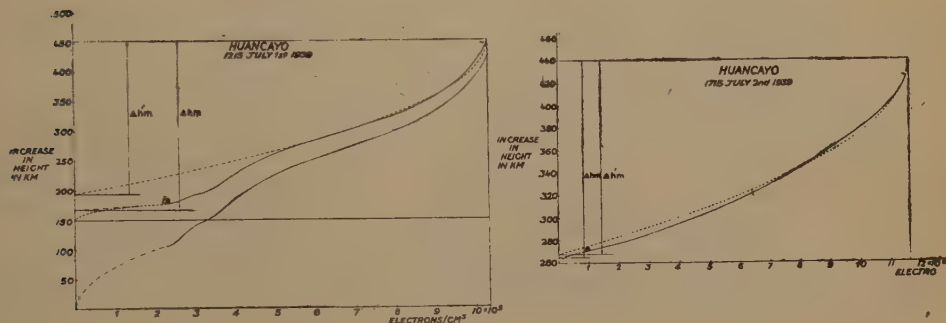
The electron density distribution three hours later on the same day is shown by fig. 3. Although the maximum electron density is somewhat

higher than before, the total number of electrons actually is 78 per cent. of its earlier value.

Fig. 4 shows two typical electron density distributions of April 4th in the same year. The distributions are parabolic over a fairly wide range and are essentially similar. Although the maximum electron density increases about 23 per cent. from 13.15 to 17.15, the total number of electrons actually decreases by about 11 per cent. It is already evident from these examples that the description of the state of the F-layer is far from complete if only the virtual heights and the maximum electron density are mentioned.

The July electron density distributions are somewhat different. Fig. 5 shows two distribution curves for July 1st at 12.15. In order to indicate the error connected with the arbitrary extension of the virtual height curve down to zero frequency, the distribution was plotted for two very different cases. The first, and most probable one, was obtained when the virtual height curve was extended by its tangent at the uncertainty

Fig. 5.



point down to zero frequency. The second and most unlikely case was obtained when the lowest point of the curve was smoothly extended by a straight line to zero height at zero frequency. It is clearly shown that the main character of the distribution is not affected much by the choice of the extension of the curve. This is, of course, also evident from the property of the Schlömilch integral equation, which represents the solution. It can, therefore, safely be stated that the electron density distribution is notably parabolic through the high density range.

In the late evening the following day the distribution is closely parabolic throughout the range. Pretty examples of similar nature are common in northern latitudes. The smooth and parabolic shape is almost indicated by the height curves shown in fig. 6.

Finally, let us take a look at two October distributions obtained the same day. They are quite similar to the April distributions just shown. Although the maximum electron density increases by about 12 per cent. from 15.15 to 18.15, the total number of electrons actually decreases much more, viz., by about 38 per cent. This shows even more clearly

the necessity of stating not only the critical frequency but also the total number of electrons of the layer (naturally only the number of electrons below the maximum density level, as the other part of the layer cannot be explored by the methods used).

The assumption of a parabolic electron distribution is especially convenient, as was pointed out by Booker and Seaton. For a characteristic

Fig. 6.

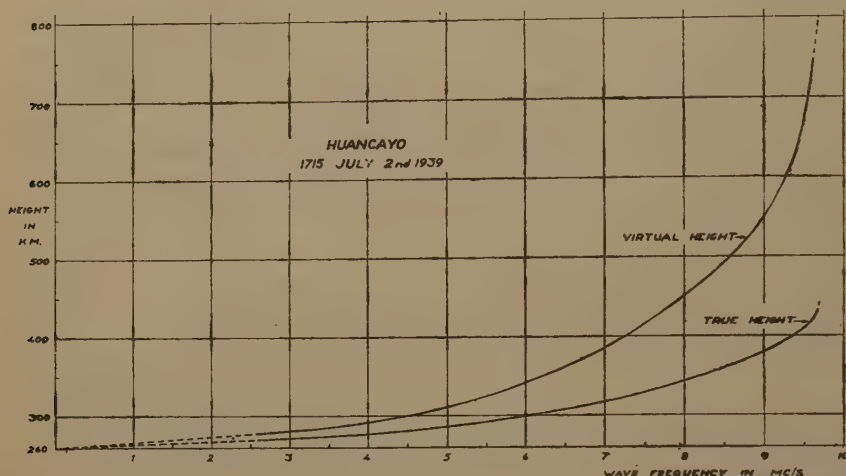
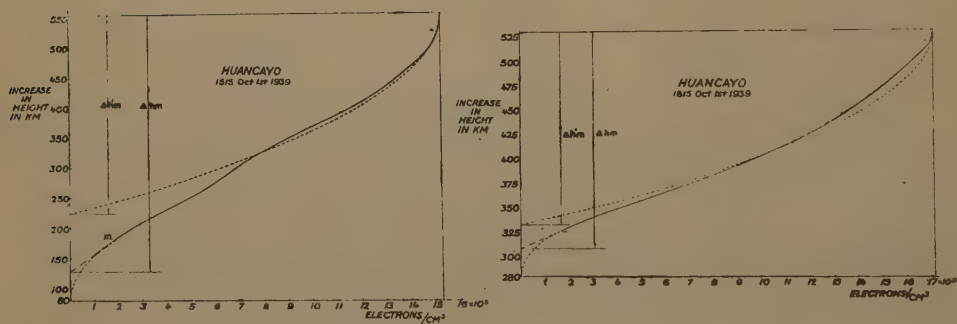


Fig. 7.



frequency, 0.834 times the critical frequency f_c of the layer, the virtual height equals the true height at the level of maximum electron density. For an Epstein layer, on the other hand, the corresponding characteristic frequency is $0.707 \cdot f_c$. The distributions of the F-layer are generally something between the parabolic and the Epstein type, as indicated by the shape and the characteristic frequency. In the following table the characteristic frequency f_k and several other characteristics of the distributions already shown are collected for the sake of convenience:—

Huancayo, 1939.

Day.	Local time.	N_t .	N_{\max} .	Decrease in N_t .	Decrease in N_{\max} .	(f_k/f_c) .	H km.
Jan. 2	15.15	$2.50 \cdot 10^8$	$2.06 \cdot 10^6$	per cent. 0	per cent. 0	0.583	76
Jan. 2	18.15	$1.97 \cdot 10^8$	$2.08 \cdot 10^6$	22	-1	0.598	51
Apr. 4	13.15	$2.19 \cdot 10^8$	$1.45 \cdot 10^6$	0	0	0.777	105
Apr. 4	17.15	$1.96 \cdot 10^8$	$1.79 \cdot 10^6$	11	-23	0.806	76
July 1	12.15	—	—	—	—	0.812	115
July 2	17.15	—	—	—	—	0.802	86
Oct. 1	15.15	$3.68 \cdot 10^8$	$1.53 \cdot 10^6$	0	0	0.746	165
Oct. 1	18.15	$2.30 \cdot 10^8$	$1.71 \cdot 10^6$	38	-12	0.786	99

N_t denotes the total number of electrons in km.electrons.cm.⁻³ in a column of 1 cm.² cross-section reaching from the bottom of the layer to the level of max. electron density, N_{\max} .

(f_k/f_c) always increases towards evening, indicating that the layer becomes more parabolic. Generally (f_k/f_c) lies somewhere between the parabolic layer value and the Epstein layer value. The H value is very approximate, so great values like 165 km., for example, should not be taken too seriously.

Finally, it should be stressed again how necessary it is to give an account of the total number of electrons. The change in critical frequency very often is entirely misleading, and only the change in the total number of electrons gives an indication of the actual change in the layer. Characteristic, for example, is that on October 1st, 18.15, the maximum electron density increased by about 12 per cent. counted from 15.15, whereas the total number of electrons actually decreased by about 38 per cent. in the same time. The assumption of a day-time expansion of the F₂-atmosphere is very well justified by these values.

For comparison a few results from Cambridge, Mass., obtained by the author should be shown.

Fig. 8 shows the electron density distribution obtained late on a quiet April afternoon. The F₁- and F₂-layers are just merging. (f_k/f_c) is 0.745 and the layer appears somewhat parabolic in its upper part.

In the small and early hours of the morning the quiet night distributions were often practically parabolic. Two typical examples are shown in fig. 9 and fig. 10.

Fig. 8.

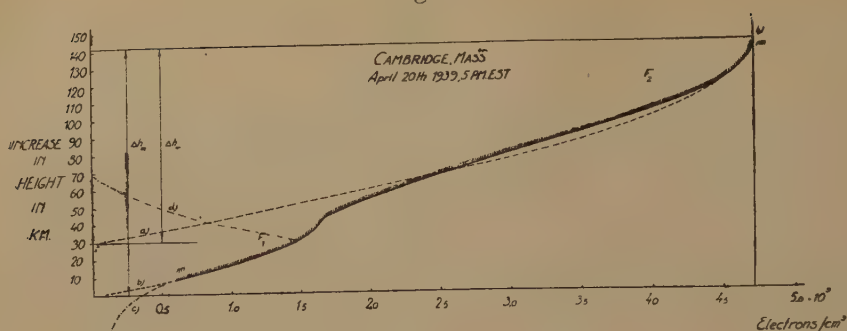


Fig. 9.

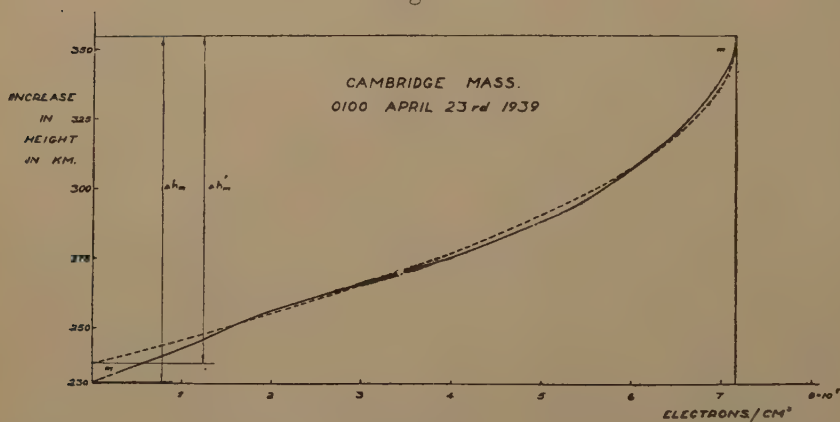
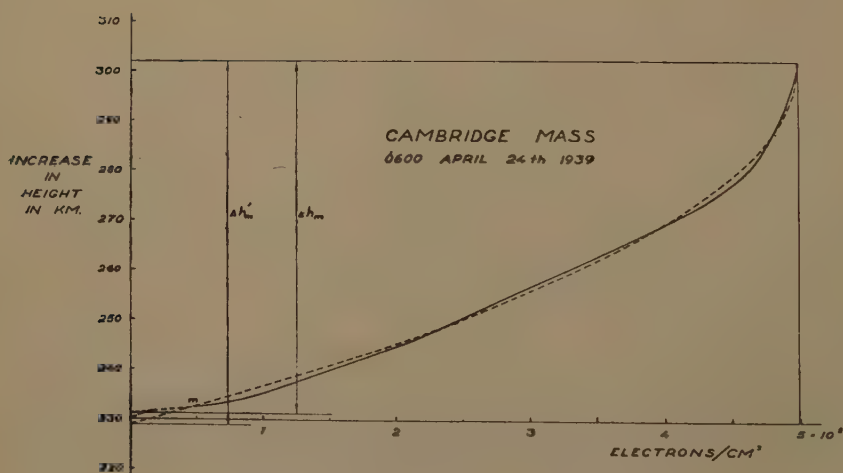


Fig. 10.

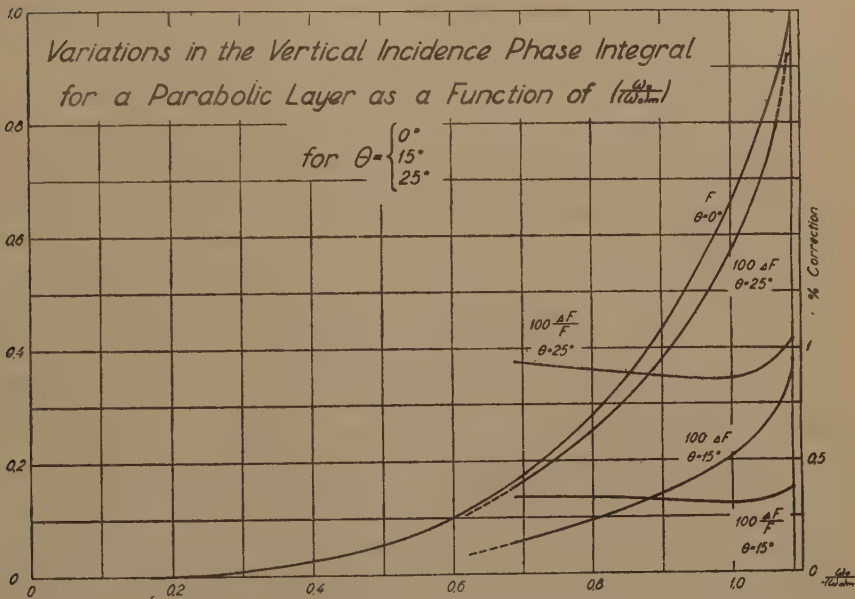


The corresponding (f_k/f_c) values are 0.843 and 0.822 respectively. H is 58 and 37 km. respectively. These values are fairly normal.

The Cambridge electron density distributions were calculated under the assumption that the angle θ between the ray direction (the vertical) and the vector of the terrestrial magnetic field is zero. In an earlier communication ⁽¹⁾ the corresponding phase integral errors were conservatively estimated. Since then a computation of the actual error for a parabolic layer has been made for propagation angles of 15° and 25° .

Even when θ is as great as 25° the relative error in the phase integral, dF/F , is only as low as about 1 per cent. From the phase correction data the error in the virtual height $d(\Delta h_v)/\Delta h_v$ is easily constructed. The result is shown in fig. 12. The correction in the virtual height is even smaller than the phase correction. Even for angles as great as 25° it is

Fig. 11.

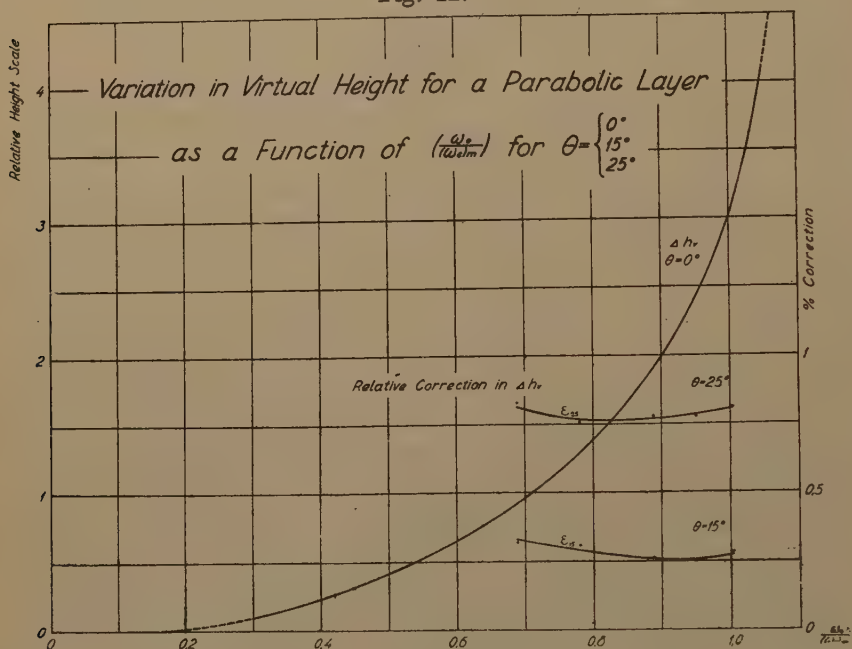


easily seen that the correction is small compared to the experimental errors. The method developed by the author ⁽¹⁾, *loc. cit.*, can therefore be used with sufficient accuracy by ionospheric observatories in such locations where θ is smaller than about 25° . The method, therefore, should generally be quite useful. So much more so as one very often finds that (as has also been shown by the results in this communication) the layer shape deviates quite a bit from the ideal parabola. To use a characteristic frequency such as $0.834 f_c$ for the routine scaling of ionosphere records, as suggested by Booker and Seaton, does not seem to be a permissible approximation in most cases. The actual (f_k/f_c) value is generally appreciably lower, and its value varies with the latitude as well as with the season and the day.

The accurate solution of the integral equation can be obtained quite fast, once one gets used to the procedure, and it furthermore seems quite reasonable that a useful, not too complicated machine could be developed to solve the integral equation, or rather to draw the true height curve direct from the experimentally obtained virtual height data. With or without a machine the integral equation should be used, as this is the only way to obtain a reliable result.

It should be added that the exact treatment of the wave equation leads to a finite virtual height at the penetration frequency. This height,

Fig. 12.



however, is very great and it is practically impossible to note the difference between the classical and exact heights on the ionosphere record unless the transmitter is very powerful.

Furthermore, the discrepancy region is so narrow, generally only a few thousand cycles wide, that we are entirely justified to use the methods which have led to the results just shown.

The author wishes to express his thanks to Alice and Knut Wallenbergs, Stiftelse (Foundation), Stockholm, the support of which has made the presentation of these results possible.

REFERENCES.

- (1) O. Rydbeck, "Jonosfärens Fysik och dess Betydelse för Radiokommunikationerna," *Tekniska Samfundets Handlingar*, No. 3, Gothenburg, Sweden (1939); "The Propagation of Electromagnetic Waves in an Ionized

- Medium and the Calculation of the True Heights of the Ionized Layers of the Atmosphere," *Phil. Mag.* ser. 7, vol. xxx. p. 282 (October 1940); "A Theoretical Survey of the Possibilities of Determining the Distribution of the Free Electrons in the Upper Atmosphere," *Transactions of Chalmers University*, No. 3, Gothenburg, Sweden (1942).
- (2) H. G. Booker and S. L. Seaton, "Relation Between Actual and Virtual Ionospheric Height," *Phys. Rev.* vol. lvii. No. 2, p. 87 (1940); K. Rawer, "Zür Frage der Auswertung von Ionosphärenbeobachtungen," *Hochfrequenztechnik und Elektroakustik*, Heft 3, Band 58 (Sept. 1941).

XVIII. *The Symmetrical Vortex Street in Sound-Sensitive Plane Jets* *.

By P. SAVIC, Ph.D. and J. W. MURPHY †.

[Received October 6, 1942.]

IN a previous paper⁽¹⁾ one of the authors has applied the Tollmien-Hollingdale criterion⁽²⁾ of hydrodynamical instability to the problem of sensitive gaseous jets and edge tones. It could be shown that the assumption of neutral oscillations, *i. e.*, periodic motion of the jet without increment or decrement, led to a solution of the hydrodynamical disturbance equation which was in good agreement with Brown's⁽³⁾ experimental results on edge tones. It was concluded that once the initial periodic disturbance is set up in the jet, the motion develops in the sense of the Tollmien-Hollingdale criterion for instability.

The solution referred to in the previous paper⁽¹⁾ corresponds to a vortex arrangement of the Kármán Street type. In the well-known paper by Kármán and Rubach⁽⁴⁾, it is shown that two parallel rows of equally spaced vortices can only exist in two arrangements, *i. e.* either symmetrical or alternate. It is further shown that in an ideal fluid the symmetrical vortex street is unstable and tends to rearrange itself in the alternate position, and that, moreover, the alternate position is only stable for one particular ratio between the longitudinal and transversal spacing of the vortices. This arrangement is the so-called Kármán Vortex Street in an ideal fluid.

In the Tollmien-Hollingdale theory the alternate vortex street appears as a consequence of the general considerations of instability. It was previously shown⁽¹⁾ that this treatment does not yield a condition for the spacing ratio of the vortices. It can merely be said that if the velocity profile of a plane jet, in the mathematical form derived by Bickley⁽⁵⁾, be inserted into the Tollmien-Hollingdale disturbance equation, then an alternate vortex street results and the "wave length" (*i. e.*

* Amplified section of a Ph.D. Thesis approved by the University of London.

† Communicated by the Authors.

the longitudinal spacing of the vortices) as well as the "wave velocity" of the vortex system can be predicted.

It was, however, observed that besides the ordinary alternating vortex street a symmetrical arrangement of the vortices can be produced under certain conditions. Rosenbrook⁽⁶⁾, who investigated the flow in a two-dimensional channel with slightly diverging walls, stresses the fact that, besides the alternate vortex street, symmetrical oscillations can sometimes be observed. The difference between the two types of flow lies in the nature of the amplitude function of the oscillations. The alternate position of the vortices corresponds to an even amplitude function, the symmetrical position to an odd one. Rosenbrook refers to Tollmien's original paper⁽²⁾, in which was stated that symmetrical oscillations "do not generally appear so that we need not pursue this possibility any further." Hollingdale, investigating the flow in the wake of a solid body, showed that if the velocity profile be represented by a cosine function, then the odd solution would give an imaginary eigenvalue for the wave-length. He concludes that "the method of proof suggests that this conclusion will apply to profiles in general" *.

It must thus be admitted that a rigorous proof of the impossibility of symmetrical oscillations does not exist. On the other hand, Rosenbrook's observations, as well as photographs of symmetrical vortex streets in sensitive gaseous jets taken by Brown⁽⁷⁾, seem to suggest that symmetrical oscillations are in fact physically possible and should be obtainable from the general disturbance equation. In other words, the odd eigenfunction satisfying both the disturbance equation and the boundary conditions should give a real value for the wave-length.

We proceed exactly as in the previous paper. Bickley's velocity profile of a plane jet is inserted into the disturbance equation, which gives equ. (7) of ref. ⁽¹⁾. We shall, however, now write this equation in non-dimensional units, thus :

$$\phi'' - \alpha^2 \phi + 6\phi \cdot \text{sech}^2 y = 0. \quad \dots \dots (1)$$

ϕ denotes the amplitude eigenfunction of the oscillations and y is the co-ordinate at right-angles to the main stream. The eigenvalue is given

by $\alpha = \frac{2\pi}{\lambda}$, where λ is the wave-length (longitudinal spacing of the vortices).

We shall formulate the boundary conditions so as to obtain a symmetrical vortex street. This requires the following restrictions upon ϕ :

ϕ an odd function of y ,

$\phi \rightarrow 0$ as $y \rightarrow \infty$.

* Dr. Hollingdale has kindly informed us that this proof was by no means meant to be rigorous, and that exceptions to general physical considerations like the following solution are conceivable.

The solution satisfying both (1) and the boundary condition is

$$\phi = k \cdot \sinh y \cdot \operatorname{sech}^2 y, \quad \dots \dots \dots (2)$$

giving the eigenvalue

$$x = 1. \quad \dots \dots \dots (2a)$$

Thus the eigenvalue is real and the solution is physically significant.

The total streamfunction, for an observer at rest relative to the orifice is

$$\Psi_1 = \tanh y + k \sinh y \cdot \operatorname{sech}^2 y \cdot \cos x. \quad \dots \dots \dots (3)$$

For an observer moving with velocity $c=2/3$ in the direction of the motion of the vortex street,

$$\Psi_2 = -\frac{2y}{3} + \tanh y + k \sinh y \cdot \operatorname{sech}^2 y \cdot \cos x. \quad \dots \dots \dots (4)$$

Fig. 1.

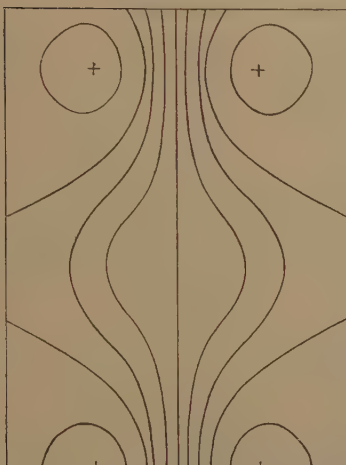
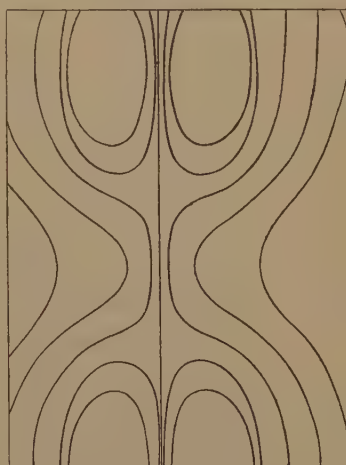


Fig. 2.



It should be noted that (3) is the real part of the complex stream function for the particular time $t=0$. In general this function is time-dependent. It was shown in the previous paper that the velocity $c=2/3$ is the speed with which the vortex pattern is propagated. The stream function (4) is therefore stationary and corresponds to the paths traced by particles suspended in the stream. The arbitrary constant k is bound to appear, owing to the homogeneity of (1). The only restriction imposed upon it is that it be small compared with unity in order that the velocity of the disturbance may be small compared with that of the main stream.

The streamlines of (3) are represented in fig. 1, the ones of (4) in fig. 2. It can be seen that fig. 2 corresponds in a general manner to Brown's photographs (*e. g.* fig. 4, plate 1 of ref. (7)). It should be noted that the increment of the oscillations is very much smaller than in the case of asymmetrical oscillations.

We shall now derive an expression for the spacing ratio of the vortices.

The longitudinal spacing is already given by the eigenvalue and can be written as

$$\lambda = 2\pi. \quad (5)$$

To find the transversal spacing we proceed as follows:—It can be seen from (3) that the streamline field is symmetrical with respect to the lines $\cos x = 1$. The centres of the vortices therefore lie on these lines. We shall adopt the following notation for the sake of brevity:

$$\tanh y = z.$$

Hence (3) can now be written thus:

$$k^2 z^4 + z^2(1 - k^2) - 2\Psi_1 z + \Psi_1^2 = 0, \quad (6)$$

where $\cos x$ has been equated to unity. This is a fourth degree equation in z , and the problem to find the vortex centres will consist in finding a value for Ψ_1 , for which (6) will have two coinciding real roots and two conjugated complex roots. We shall therefore attempt to factorize (6) by the expression:

$$z^2 - 2Az + A^2, \quad (7)$$

having the two coinciding real roots $z_1 = z_2 = A$. Dividing (6) by (7) gives

$$\frac{k^2 z^4 + z^2(1 - k^2) - 2\Psi_1 z + \Psi_1^2}{z^2 - 2Az + A^2} = k^2 z^2 + 2Ak^2 z + 1 - k^2 + 3A^2 k^2 + R.$$

R denotes the remainder of the division, and the condition that (7) is a factor of (6) can only be fulfilled if $R = 0$. A simple calculation shows that this is the case if

$$\Psi_1 + A^3 k^2 = A - Ak^2 + 3A^3 k^2,$$

$$A^2 - A^2 k^2 + 3A^4 k^2 = \Psi_1^2;$$

after eliminating Ψ_1 ,

$$A = \tanh y_0 = \pm \frac{\sqrt{2}}{4k} \sqrt{-1 + 4k^2 + \sqrt{1 + 8k^2}}.$$

y_0 is now the distance of the vortex centre from the line $y = 0$. The transversal spacing of the vortices is therefore given by $h = 2y_0$.

This gives, together with (5), the final expression for the spacing ratio:

$$\frac{h}{\lambda} = \frac{1}{\pi} \tanh^{-1} \left(\frac{\sqrt{2}}{4k} \sqrt{-1 + 4k^2 + \sqrt{1 + 8k^2}} \right). \quad (8)$$

Exactly as in the previous case of asymmetrical oscillations, it can be seen that the spacing ratio depends on the arbitrary constant k . It may be asked what the value of k must be in order to make (8) equal to the Kármán spacing ratio. Kármán has shown that for the stable alternate position the spacing ratio is given by

$$\frac{h}{\lambda} = \frac{1}{\pi} \tanh^{-1} 1/\sqrt{2}.$$

Comparing this with our expression (8), it follows that Kármán's ratio can only be realized if k be taken as zero. Then, however, the whole

vortex term in our stream function (3) must vanish as well and the motion becomes stationary. This means that it is impossible for the symmetrical vortex street to assume the Kármán ratio. This result is hardly surprising in view of the fact that Kármán showed that all symmetrical vortex streets are unstable. It requires the introduction of viscosity to derive the Tollmien-Hollingdale criterion, and the results derived from it must needs contradict Kármán's considerations of an ideal fluid.

A further point to be investigated is, whether numerical agreement with experiment can be achieved. The equations (13) and (14) of the previous paper (ref. ⁽¹⁾) were derived from the two eigenvalues λ and c . Exactly the same thing can be done with the present odd solution, and since (5) is exactly twice the value of the corresponding even solution (equ. (9) of ref. ⁽¹⁾), the equations for λ and c/v remain unaltered except for the constant which becomes halved, thus :

$$\text{and } \left. \begin{aligned} \frac{c}{v} &= \left(\frac{m \cdot n}{0.477 \dots v} \right)^{1/3} \\ \lambda &= \left(\frac{m \cdot v^2}{0.477 \dots n^2} \right)^{1/3} \end{aligned} \right\} \dots \dots \dots (9)$$

TABLE I.

n .	m .	v .	λ exp.	λ theor.
97	0.4	85.6	0.80	0.87
97	0.4	96.2	0.86	0.94
97	0.3	74.0	0.62	0.79
97	0.3	89.0	0.77	0.81

Here m again denotes the width of the orifice of the jet, n the frequency of the oscillations and v the stream velocity at the orifice. Comparison with Brown's results on sensitive jets (table 2 of ref. ⁽⁷⁾) is provided in our Table I. The agreement is on the whole bad. This can probably be ascribed to the fact that the oscillations begin at a point very close to the orifice. It must be assumed that Bickley's velocity profile, which was derived from Prandtl's boundary layer equations, does not apply here. Moreover, the disturbance by sound is probably of the order of magnitude of the stream velocity, so that the approximation involved in the disturbance equation is insufficient. An improvement of the theory should therefore take account of these points.

Summarizing, it can now be said that, according to the Tollmien-Hollingdale disturbance equation, two types of neutral oscillation can exist in a plane jet. Although the criterion applies to spontaneous instability, the previous investigation has shown that the fact that instability produced by sound is responsible for asymmetrical oscillations does not upset the argument, provided the displacement velocity of the

sound is small compared with the velocity of the main stream. If the stream velocity is small the symmetrical type of oscillation is set up, but here the displacement velocity of the sound is probably of the order of the velocity of the main stream, and the agreement with experiment is accordingly not good.

According to the theory advanced in this and the previous paper, edge tones are produced by the mechanism proposed by Brown⁽⁸⁾, but once the initial disturbance is set up, the motion obeys the laws of the Tollmien-Hollingdale criterion. Similarly, the organ- or flue-pipe works by virtue of an interaction between the free edge tone and the resonating column of air. Brown⁽⁹⁾ has shown that the mechanism of the flue pipe can be understood entirely on the grounds of this interaction. The edge tone can be produced in different "stages."

The state of an edge tone system is said to be in stage 1 if the distance between orifice and edge lies between one and two wave-lengths of the vortex motion, it is in stage 2 if this distance lies between two and three wave-lengths, and so forth. If the "normal" organ-pipe tone is brought about by stage 2, then overblowing corresponds to the occurrence of stage 3, whereas underblowing is equivalent to the appearance of stage 1. If, however, stage 1 is responsible for the normal tone, then it is difficult to see how underblowing is produced. The present solution offers a possibility of understanding this feature. As the velocity of the air blast in a flue-pipe decreases, the Reynolds number of the jet becomes smaller than the critical Reynolds number for production of asymmetrical oscillations, but is still large enough for the production of symmetrical oscillations. The former type will therefore disappear and will be replaced by the latter. This, then, is the underblown tone. It can be prevented by a "beard" in the form of a lateral obstruction of the jet. An asymmetry being thus produced in the jet, the production of the symmetrical vortex street is suppressed.

It is intended to deal with experimental evidence of the detailed mechanism of sound-sensitive jets in a future communication.

We are indebted to Dr. G. B. Brown for supplying the values for Table I. and for valuable help, and also to Dr. S. H. Hollingdale and Mr. H. B. Squire, M.A., for discussion and criticism.

References.

- (1) Savic, *Phil. Mag.* xxxii. p. 245 (1941).
- (2) Tollmien, *Nachr. Ges. Wiss. Goettingen*, p. 79 (1935); Hollingdale, *Phil. Mag.* xxix. p. 209.
- (3) Brown, *Proc. Phys. Soc.* xlix. p. 493 (1937).
- (4) Kármán & Rubach, *Phys. Zeitschr.* xiii. p. 48 (1912).
- (5) Bickley, *Phil. Mag.* xxiii. p. 727 (1937).
- (6) Rosenbrook, *Zeitschr. angew. Math. Mech.* xvii. p. 8 (1937).
- (7) Brown, *Proc. Phys. Soc.* xlvii. p. 703 (1935).
- (8) Brown, *Proc. Phys. Soc.* xlix. p. 508 (1937).
- (9) Brown, 'Nature,' cxli. p. 11 (1938).

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

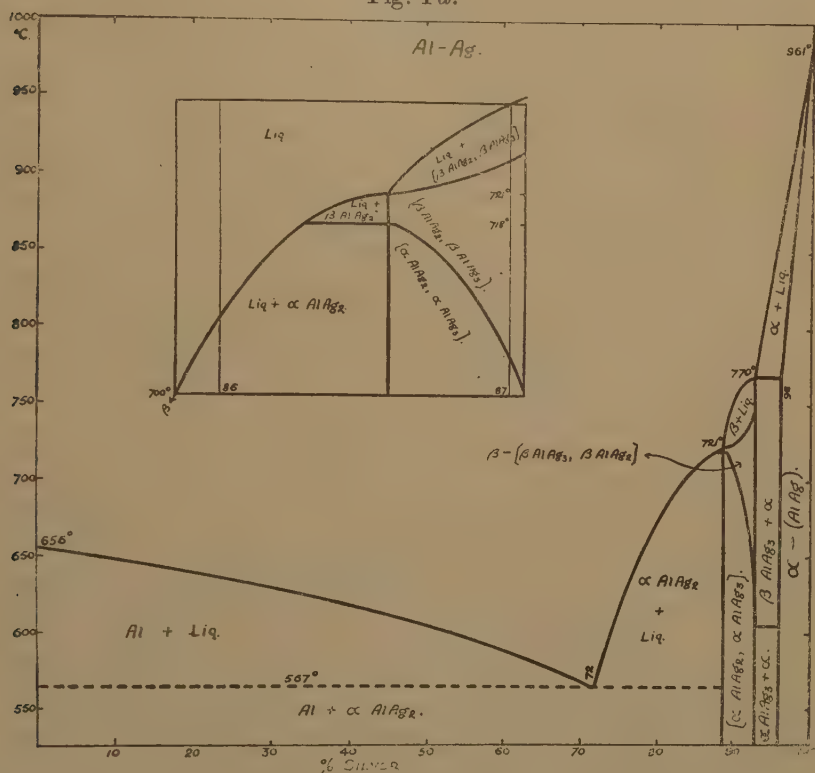
XIX. *The Hall Effect and some other Physical Constants of Alloys.*—
Part VII. *The Aluminium-Silver Series of Alloys.*

By HANDEL POWELL, Ph.D., and Prof. E. J. EVANS, D.Sc.,
Physics Department, University College of Swansea *.

[Received October 8, 1942.]

ACCORDING to the equilibrium diagram of the aluminium-silver system arrived at by Petrenko ⁽¹⁾ and reproduced in fig. 1 *a*, there should be two

Fig. 1 *a*.



intermediate phases present at ordinary temperatures: one corresponding to Ag_3Al and the other being a mixed crystal phase made up of Ag_3Al and Ag_2Al . Westgren and Bradley ⁽²⁾, as the result of an X-ray investigation of the structure of these alloys, confirmed in general the deductions of Petrenko from thermal data. According to their observations the α phase, which is a solid solution of aluminium in silver, extends

* Communicated by the Authors.

from pure silver to a composition containing 5.9 per cent. by weight of aluminium, and the lattice constant falls in value from 4.079 (pure Ag) to 4.053 Å. The intermediate β' phase was found to correspond exactly to the formula Ag_3Al (7.73 per cent. Al), with a narrow range of homogeneity. Its structure was cubic, and the unit cell having an edge of 6.920 Å contained 20 atoms. The other intermediate phase (γ) of the system is homogeneous in a range extending from 8.5 to 14.35 per cent. by weight of aluminium, and is a solid solution of closely-packed hexagonal structure. The lattice dimensions of this hexagonal structure vary continuously from one boundary to the other of this homogeneous phase. From X-ray photograms it was concluded that there was no reason to believe that the γ phase contains an intermetallic compound with a formula such as Ag_2Al or Ag_3Al_2 .

The β' phase was found to be in equilibrium with α in the range 5.9 to 7.73 per cent. by weight of aluminium, and with γ in the range 7.73 to 8.5 per cent. by weight of aluminium. The X-ray measurements showed that the solubility of silver in aluminium is very small, and that the range of composition from 14.35 to 100 per cent. by weight of aluminium is a mixture of aluminium and the γ phase. The α solid solution of aluminium in silver has been the subject of an X-ray investigation by Phelps and Davey⁽³⁾, who found that saturation was reached when 5.4 per cent. of aluminium by weight was added to silver.

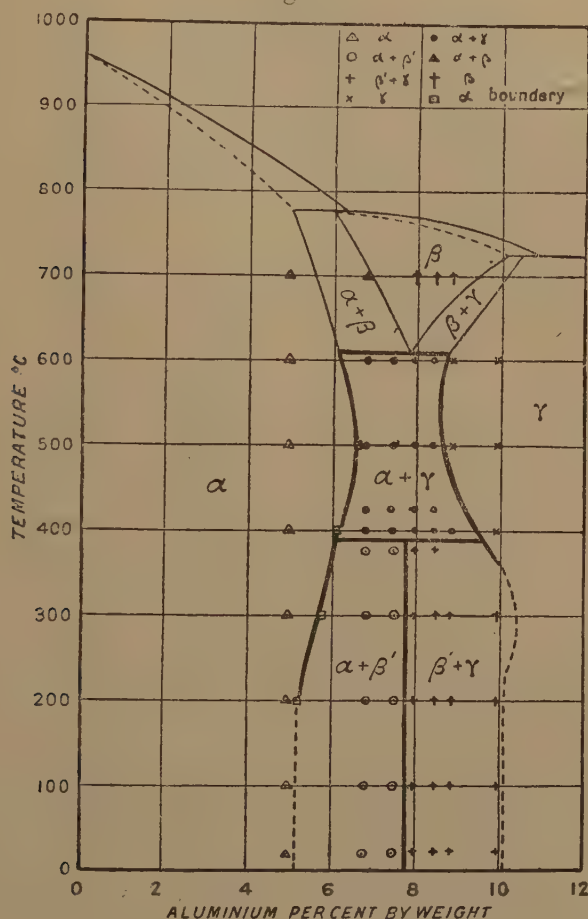
The equilibrium diagram for the silver-rich alloys for temperatures above 600° C. has been established with great accuracy by Hoar and Rowntree⁽⁴⁾ by means of thermal and micrographic investigations. The silver end of the aluminium-silver has also been investigated by Ageew and Shoyket⁽⁵⁾. The equilibrium diagram deduced from their X-ray, micrographic, and hardness observations is reproduced in fig. 1*b*. They determined the boundary of the α phase and found that it corresponds to a composition of 6.52 per cent. by weight of aluminium at 500° C. and to 5.10 per cent. by weight of aluminium at 200° C. The β phase is stable only at high temperatures, and on cooling decomposes into an eutectoid mixture, which is converted into a new β' phase (Ag_3Al) at 400° C. The ($\alpha+\beta'$) phase extends from about 5 per cent. by weight to 7.73 per cent. by weight of aluminium, and the ($\beta'+\gamma$) phase from 7.73 per cent. by weight to about 10.1 per cent. by weight of aluminium.

Crepaz⁽⁶⁾ has made a thermal analysis and a micrographic examination of the aluminium-silver system. His results indicate the presence of the three compounds Ag_3Al , Ag_2Al , and Ag_3Al_2 . These form limited series of solid solutions with one another, with aluminium, and with silver. The eutectic between Ag_3Al_2 and aluminium saturated with silver contains 39.5 per cent. of silver and melts at 567° C. The solid solubility of silver in aluminium falls from about 45 per cent. at the eutectic temperature to less than 0.5 per cent. at room temperature.

Measurements of hardness, electrical conductivity, and temperature-coefficient of resistance carried out by Le Grix and Broniewski⁽⁷⁾ indicated the presence of the compounds Ag_3Al_2 and Ag_3Al .

In the present investigation sixteen alloys of various compositions were carefully prepared from pure aluminium and pure silver, and their resistivities, temperature coefficients of resistance, thermoelectric powers, and Hall coefficients were measured after annealing at temperatures deduced from the equilibrium diagram. The experimental results

Fig. 1 b.



obtained will be discussed later in relation to thermal, micrographic and X-ray observations carried out on the system.

PREPARATION OF THE ALLOYS.

The alloys were prepared from pure silver obtained from Johnston-Matthey, and from pure electrically refined aluminium (purity 99.99 per cent.) obtained from Aluminium Industries, A. G. Neuhausen, Switzerland. The pure silver and aluminium were mixed in a salamander crucible, and melted in an electric furnace. The elements readily alloy together

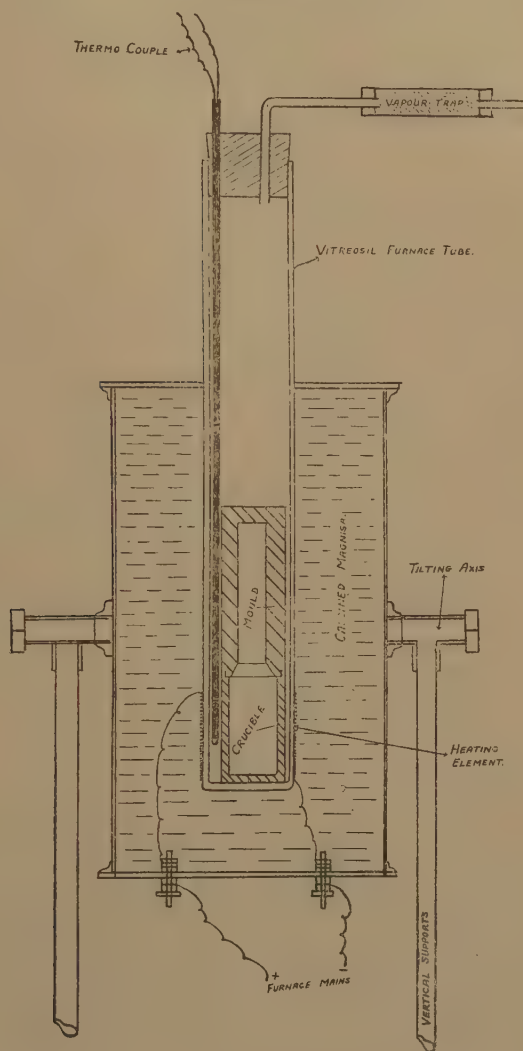
throughout the whole range of alloy compositions, and the homogeneity of the alloy under preparation was ensured by thoroughly stirring the melt with a carbon rod. The melt was poured into graphite moulds and cast into rectangular plates of dimensions 12 cm. \times 2.5 cm. \times 0.4 cm. approximately. With the exception of the high percentage silver alloys, the use of a flux to prevent oxidation was unnecessary as the aluminium readily combines with the oxygen of the atmosphere to form a tough skin of aluminium oxide on the surface, and thus protects the rest of the melt. For the high percentage silver alloys (90–100 per cent. Ag) carbon was effectively used as a flux. The mould itself was heated to a temperature of about 80° C. below the melting point of the alloy to be cast, as this procedure was found to assist the casting of satisfactory plates. The tough skin of aluminium oxide had almost invariably to be broken in order to start the flow of the alloy from crucible to mould.

As it is difficult to prepare satisfactory plates of aluminium alloys free from oxides and blowholes, other methods of preparation were tried. The first of these was to prepare and cast the alloys *in vacuo*, and the apparatus employed for this purpose is shown in fig. 2. The two metals were placed in the crucible, to the top of which was fitted the mould. The crucible and mould were placed in a vitreosil tube, closed at its lower end and wound over the lower portion with nichrome wire which constituted the heating element of the electric furnace. Through a rubber bung at the upper end of the vitreosil tube passed a tube containing a thermocouple, and another tube which could be connected to a pump. The vitreosil tube was contained in a large metal cylinder capable of rotation about a horizontal axis, and the space between the tube and cylinder was packed with calcined magnesia, which formed the lagging element of the furnace. The vitreosil tube was evacuated, and then an electric current was passed through the heating element until the molten alloy was formed. The complete mixing of the metals was ensured by quickly rotating the cylinder on its axis through an angle of about 40° on either side of the vertical. The furnace was then tipped upside down so as to allow the melt to pass from the crucible into the mould. It was found that the values of the densities of the alloys cast *in vacuo* were slightly less than those previously obtained by casting in air in the usual way. The graphite crucibles and moulds used in the above experiments were replaced by others made from finest alundum, but the change had no effect on the densities of the alloys.

The second method consisted in preparing and casting the plates in an atmosphere of pure dry nitrogen. The apparatus was very similar to that employed for casting *in vacuo*, except that the rubber bung closing the vitreosil tube carried three tubes. The first of these, as before, was connected to a vacuum pump, the second to a cylinder of nitrogen, and the third carried the thermocouple. The nitrogen was thoroughly purified and dried before entering the furnace. Various types of crucibles and moulds were used, but no increase in the values of the densities were obtained.

A method which was used with much success by Hanson and Slater⁽⁸⁾ for the preparation of sound alloys was next employed. It involved the bubbling of an inert gas through the melt before casting in order to remove the other gases, such as oxygen and hydrogen, which are occluded in the

Fig. 2.



melt. In the present experiments a mixture of nitrogen and carbon tetrachloride was bubbled through the melt for about five minutes, and the molten alloy was then thoroughly stirred by means of a carbon rod. The melt was cast by passing it through a red hot carbon funnel into a carbon mould, which was heated to a temperature of about 400°C . It was found that there was a slight improvement ($\frac{1}{3}$ to $\frac{1}{2}$ per cent.) in

the densities of the alloy plates as compared with the densities obtained by casting plates of the same composition in air in the ordinary way.

Finally, plates were prepared by the method of pressure casting. The molten alloy was allowed to solidify in a mould placed inside a strong steel cylinder, which was connected to a cylinder containing nitrogen at high pressure. Different pressures varying from 50 lbs. per sq. in. to 200 lbs. per sq. in. were used during the course of the experiments. The values of the densities of the alloys produced in this way were the same as the values obtained by casting in air in the ordinary way, but were higher than the values obtained when the same alloys were cast *in vacuo*.

After casting, the plates were smoothed down until they were nearly of uniform thickness and width, and the variation in thickness of the plates was found to be in general less than 0.5 per cent.

ANALYSIS OF THE ALLOYS.

The silver content of each alloy was estimated by dissolving a known small quantity of the alloy in nitric acid, and cautiously heating the solution to get rid of oxides of nitrogen. The solution was cooled, made up to a known volume, and titrated against a standard solution of potassium thiocyanate. The aluminium content was obtained by difference. In the case of alloys which could be easily planed, three samples from each plate were taken for analysis, one from each end of the plate and a portion of the turnings obtained by planing the plate surfaces. From the more brittle plates only two samples were taken, one from each end of the plate. The values of the silver content determined by the analysis of the specimens taken from one plate agreed to within 0.2 per cent., and consequently the plates were assumed to be of uniform composition throughout their volume. The compositions of the alloys are given in Table I.

ANNEALING OF THE ALLOYS.

The alloys were annealed in an electrically heated vacuum tubular furnace for periods ranging from twelve to twenty-four weeks. The uniform temperature at the centre of the furnace was recorded by a chromel-alumel thermocouple connected to a galvanometer. The thermocouple was calibrated by measuring the deflexions of the galvanometer corresponding to the known melting points of the pure metals, lead, tin, zinc, aluminium and silver. The resistivity of each alloy plate at 0° C. had already been determined immediately after preparation. A group of plates separated from each other by sheets of crown glass or sheets of polished fireclay were placed inside the furnace in the uniform temperature region, and annealed at the highest possible temperature, which was ascertained from the equilibrium diagram. The alloys were heated up to the annealing temperature, and maintained at that temperature from two to three weeks. The furnace was then allowed to cool slowly to room temperature, and the resistivity of each alloy determined at 0° C.

The plates were then replaced in the furnace, and the annealing at the previous annealing temperature was continued for a further period of about three weeks. After cooling to room temperature, the resistivity of each plate at 0° C. was again determined. This process of annealing was continued until no change in resistivity was obtained on further annealing, and when this occurred the plates were considered to be fully annealed. Since phase boundaries may be curved or inclined to the temperature axis, and since they are not always correctly given in equilibrium diagrams, it is difficult to be sure of the complete annealing

TABLE I.

Composition by Weight.	
Percent. of silver.	Percent. of aluminium.
0.00	100.00
10.02	89.98
30.21	69.79
50.19	49.81
70.23	29.77
80.60	19.40
84.74	15.26
87.96	12.04
88.50	11.50
90.40	9.60
91.75	8.25
92.30	7.70
94.29	5.71
95.08	4.92
96.11	3.89
97.00	3.00
99.50	0.50
100.00	0.00

of alloys situated at or near phase boundaries. In the present investigation, the alloys were therefore annealed down to room temperature, the temperature of annealing being reduced by steps varying from 200° to 50° C. from the highest annealing temperature.

After the completion of annealing, some of the plates selected at random were subjected to chemical analysis, and it was found that annealing had not appreciably affected the composition of the plates.

MEASUREMENT OF THE PHYSICAL CONSTANTS.

In a previous communication⁽⁹⁾ the methods employed in the determination of the various physical constants have been described in detail, and here it is only necessary to give a brief account of the measurements undertaken.

Density.

The alloy plates were weighed in air and in distilled water of a known temperature, and their densities determined from these measurements by means of Archimedes' principle. The densities of the alloys had been measured before annealing in order to determine the soundness of the castings, and it was found that annealing increased the densities of some of the alloys at the silver end of the series by about 0.3 per cent., so that the final values were in good agreement with those calculated from crystal structure. The densities of the alloys in the annealed state are given in Table II. and plotted in graph 1.

TABLE II.

Composition by weight.		Density in grams/c.c.
Percent. of silver.	Percent. of aluminium.	
0.00	100.00	2.71
10.02	89.98	2.90
30.21	69.79	3.45
50.19	49.81	4.30
70.23	29.77	5.72
80.60	19.40	6.75
84.74	15.26	7.30
87.96	12.04	7.85
88.50	11.50	7.97
90.40	9.60	8.39
91.75	8.25	8.80
92.30	7.70	8.95
94.29	5.71	9.08
95.08	4.92	9.30
96.11	3.89	9.49
97.00	3.00	9.75
99.50	0.50	10.30
100.00	0.00	10.49

Resistivity.

The resistivity of each alloy before and after annealing was measured at 0° C., and the experimental results are given in Table III. and plotted in graph 2.

Temperature Coefficient of Resistance.

The mean temperature coefficient of resistance of each alloy was determined over the temperature range 0–100° C., and the values are given in Table IV. and plotted in graph 3.

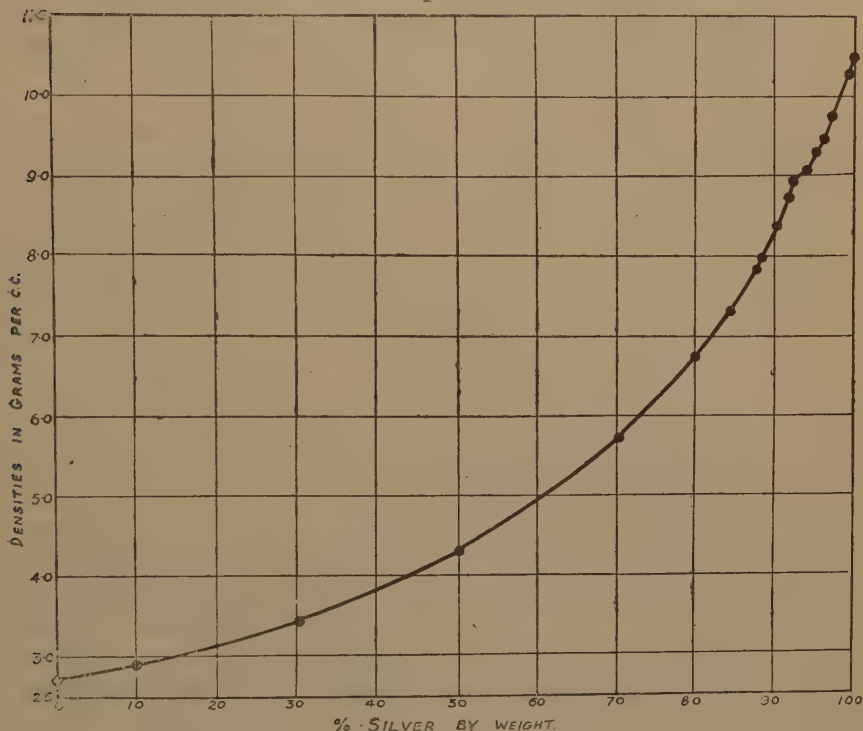
Thermoelectric Power.

The mean thermoelectric power of each alloy with respect to a plate of pure electrolytic copper was determined over the range 17° – 100° C., and the results are shown in Table V. and plotted in graph 4.

Hall Effect.

The Hall coefficient of each alloy was measured for six different fields of magnitudes 3068, 4601, 5757, 6700, 7484 and 8288 gauss. The values of the coefficients are shown in Table VI. and graph 5, whilst the relation between the Hall e.m.f. and magnetic field intensity is plotted in graph 6. These results show that, within experimental error, the Hall coefficients are independent of the magnetic field strength.

Graph 1.



DISCUSSION.

The X-ray analysis of Ag-Al system by Westgren and Bradley⁽¹⁰⁾ has shown the existence of phase boundaries at compositions corresponding to 85.6, 91.5, 92.3 and 94.1 per cent. of silver by weight. The region extending from 0 to 85.6 per cent. consists of a mixture of Al and the γ phase, and the region extending 85.6 per cent. to the silver end of the system contains four phases. The homogeneous γ phase extends from 85.6 to 91.5 per cent. Ag, the $(\beta' + \gamma)$ phase from 91.5 to 92.3 per cent. Ag, the $(\beta' + \alpha)$ phase from 92.3 to 94.1 per cent. Ag, and the α phase, which

TABLE III.

Composition by weight.		Resistivity in microhm-cm.	
Percent. of silver.	Percent. of aluminium.	Before annealing.	After annealing.
0.00	100.00	2.52	2.50
10.02	89.98	4.13	2.99
30.21	69.79	6.01	4.09
50.19	49.81	7.73	6.02
70.23	29.77	12.08	10.01
80.60	19.40	19.72	18.03
84.74	15.25	27.81	27.01
87.96	12.04	28.50	29.15
88.50	11.50	27.91	29.98
90.40	9.60	28.35	34.99
91.75	8.25	30.47	38.25
92.30	7.70	35.06	39.48
94.29	5.71	23.57	25.40
95.08	4.92	21.04	22.78
96.11	3.89	19.13	18.75
97.00	3.00	15.57	15.40
99.50	0.50	5.73	5.51
100.00	0.00	1.52	1.51

Graph 2.

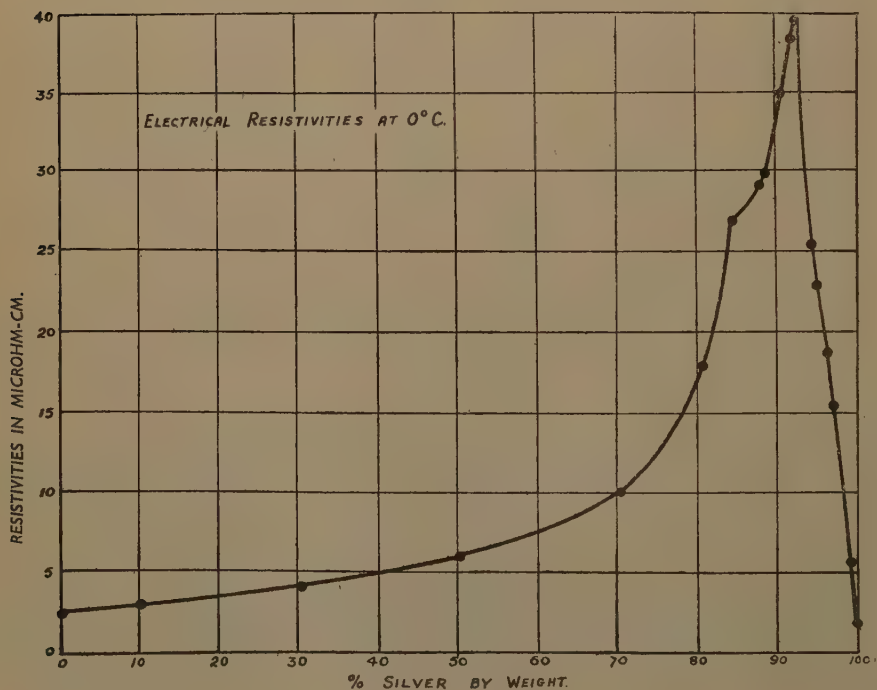
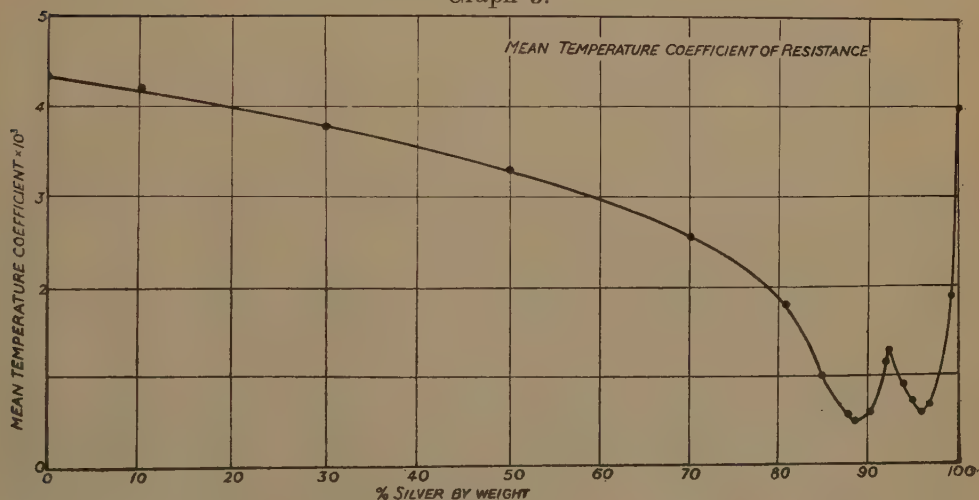


TABLE IV.

Composition by weight.		Mean temperature coefficient $\times 10^3$ between 0°C. and 100°C.
Percent. of silver.	Percent. of aluminium.	
0.00	100.00	4.35
10.02	89.98	4.21
30.21	69.79	3.79
50.19	49.81	3.30
70.23	29.77	2.54
80.60	19.40	1.80
84.74	15.26	1.00
87.96	12.04	0.57
88.50	11.50	0.49
90.40	9.60	0.58
91.75	8.25	1.15
92.30	7.70	1.29
94.29	5.71	0.91
95.08	4.92	0.71
96.11	3.89	0.59
97.00	3.00	0.67
99.50	0.50	1.90
100.00	0.00	4.00

Graph 3.



TanLE V.

Composition by weight.		Mean thermo- electric power in microvolts/°C.
Percent. of silver.	Percent. of aluminium.	
0.00	100.00	−3.25
10.02	89.98	−3.15
30.21	69.79	−2.85
50.19	49.81	−2.34
70.23	29.77	−1.29
80.60	19.40	−0.31
84.74	15.26	+0.55
87.96	12.04	−0.36
88.50	11.50	−0.51
90.40	9.60	−1.04
91.75	8.25	−3.97
92.30	7.70	−4.35
94.29	5.71	−2.50
95.08	4.92	−1.84
96.11	3.89	−0.93
97.00	3.00	−1.08
99.50	0.50	−1.40
100.00	0.00	−0.40

Graph 4.

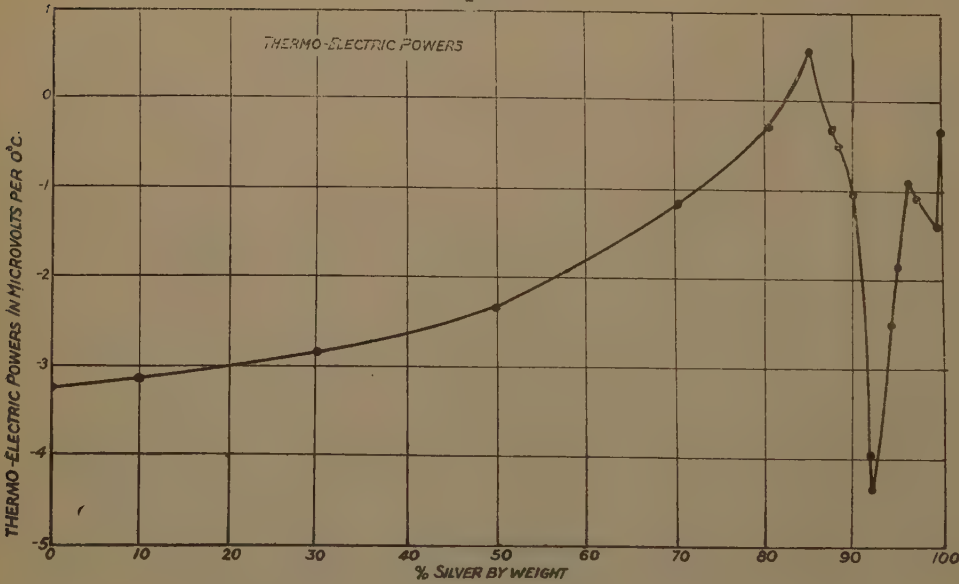


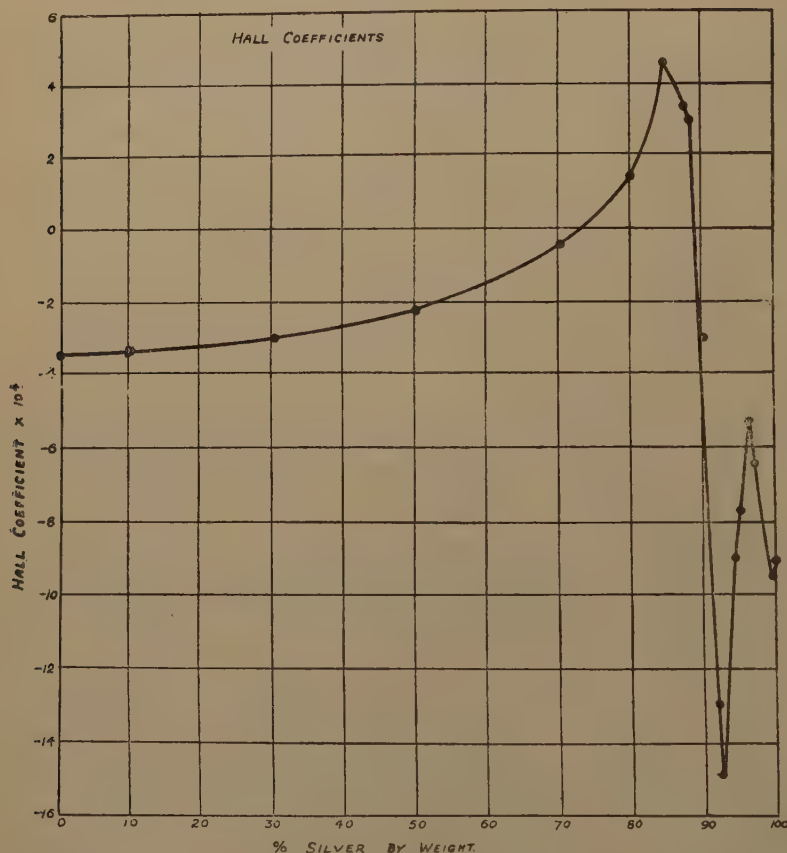
TABLE VI.
Hall Coefficients in Absolute Units.

Composition by weight.		Hall coefficient (R) $\times 10^4$.						Mean value of $R \times 10^4$.	Temp. °C.
Percent. of silver.	Percent. of Al.	Magnetic field in gauss.							
		3068.	4601.	5757.	6700.	7484.	8288.		
0.00	100.00	— 3.61	— 3.45	— 3.56	— 3.53	— 3.47	— 3.48	— 3.50	18.5
10.02	89.93	— 3.38	— 3.41	— 3.43	— 3.39	— 3.41	— 3.42	— 3.40	18.5
30.21	69.79	— 2.98	— 3.07	— 3.01	— 3.02	— 2.98	— 3.02	— 3.01	19.0
50.19	49.81	— 2.24	— 2.33	— 2.26	— 2.30	— 2.26	— 2.30	— 2.28	17.5
70.23	29.77	—	—	—	— 0.47	— 0.42	— 0.46	— 0.45	17.5
80.60	19.40	+ 1.35	+ 1.44	+ 1.37	+ 1.41	+ 1.37	+ 1.41	+ 1.39	17.5
84.74	15.26	+ 4.42	+ 4.56	+ 4.59	+ 4.47	+ 4.47	+ 4.49	+ 4.50	19.0
87.96	12.04	+ 3.27	+ 3.36	+ 3.29	+ 3.33	+ 3.29	+ 3.33	+ 3.31	19.5
88.50	11.50	+ 2.88	+ 2.99	+ 2.96	+ 2.93	+ 2.95	+ 3.01	+ 2.95	19.5
90.41	9.60	— 3.02	— 3.11	— 3.05	— 3.06	— 3.02	— 3.06	— 3.05	18.5
91.75	8.25	— 12.90	— 13.16	— 13.10	— 12.95	— 12.96	— 12.93	— 13.00	18.5
92.3	7.70	— 14.84	— 14.89	— 14.76	— 14.90	— 14.86	— 14.83	— 14.85	18.5
94.29	5.71	— 8.98	— 8.95	— 9.01	— 9.06	— 9.02	— 9.03	— 9.01	19.0
95.08	4.92	— 7.81	— 7.63	— 7.77	— 7.71	— 7.64	— 7.65	— 7.70	19.0
96.11	3.89	— 5.23	— 5.34	— 5.31	— 5.28	— 5.30	— 5.36	— 5.30	19.0
97.00	3.00	— 6.46	— 6.42	— 6.40	— 6.47	— 6.46	— 6.47	— 6.45	19.0
99.50	0.50	— 9.46	— 9.53	— 9.59	— 9.46	— 9.47	— 9.51	— 9.50	18.5
100.00	0.00	— 8.99	— 9.02	— 8.96	— 9.07	— 9.02	— 9.05	— 9.02	18.5

is a solid solution of Al in Ag, from 94.1 per cent. Ag to the silver end of the system. The β' phase was found to have a very narrow range of homogeneity and to correspond exactly to the compound Ag_3Al . One of the boundaries of the homogeneous γ phase, which is a solid solution, has a composition (85.6 per cent. Ag) corresponding to the formula Ag_3Al_2 .

According to the equilibrium diagram (fig. 1) given by Petrenko, the phase boundaries occur at compositions corresponding to 88.8, 92.3 and

Graph 5.

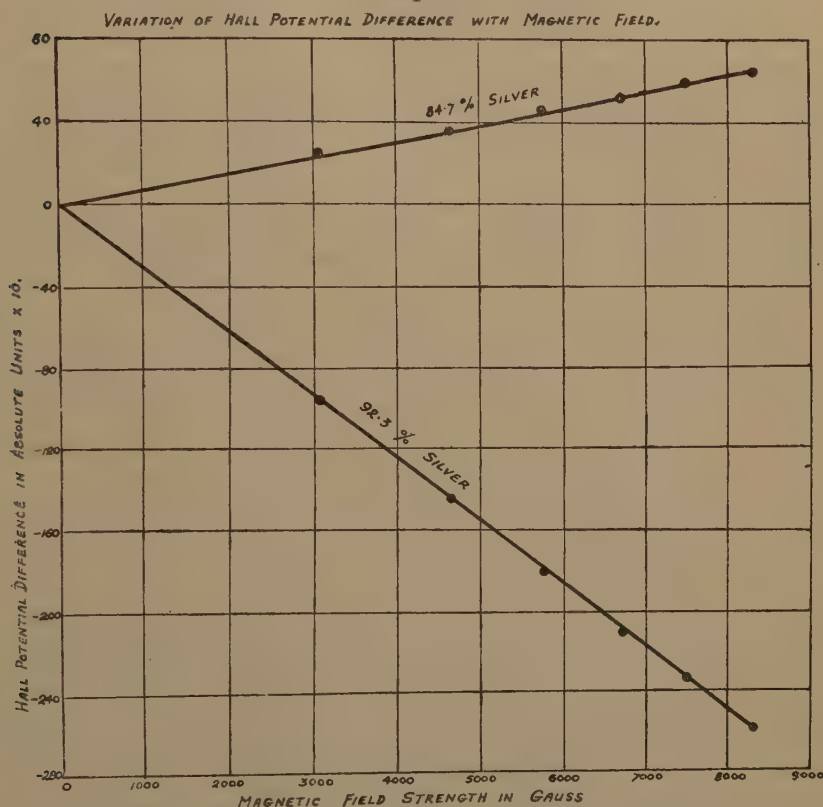


96 per cent. Ag by weight. Crepaz ⁽¹¹⁾ considered that his results indicated the presence in the system of the three compounds Ag_3Al_2 , Ag_2Al , and Ag_3Al , and Le Grix and Broniewski ⁽¹²⁾, from measurements of the physical properties of the alloys, found evidence for the existence of the compounds Ag_3Al_2 and Ag_3Al .

The physical properties of the alloys vary considerably in passing through the various phases, and the results obtained in the present investigation will now be examined in relation to our knowledge of the system as deduced by X-ray and thermal methods.

The curve (graph 1) showing the relation between density and composition shows a slight break at the composition corresponding to 92.3 per cent. Ag (Ag_3Al), and supports the conclusion that Ag_3Al is probably the only intermetallic compound in the system. The densities of the alloys in the region between 85 and 100 per cent. Ag are in reasonably good agreement with those obtained by Westgren and Bradley. The densities of the alloys in the $(\text{Al}+\gamma)$ phase are below points situated on a straight line joining the density of Al to that of the alloy containing 85.6 per cent. Ag, but the values of the densities of the alloys containing

Graph 6.



5 and 80 per cent Ag are in fairly good agreement with those given in Mellor's 'Treatise on Inorganic and Theoretical Chemistry' ⁽¹³⁾.

The variation of specific resistance with composition is given in graph 2. As the concentration of silver in the alloy increases, the curve rises gradually at first, and then more rapidly in the neighbourhood of the composition 85 per cent. Ag, which corresponds approximately to the compound Ag_3Al_2 and is the boundary between the $(\text{Al}+\gamma)$ and γ phases. The resistivity continues to increase, and the curve shows a change of slope at the composition 88.5 per cent. Ag, which corresponds approximately to the formula Ag_2Al . The curve then rises steeply to a

maximum at 92.3 per cent. Ag, which corresponds to Ag_3Al . As the resistivity decreases very rapidly for alloys of greater Ag concentration, the peak at 92.3 per cent. Ag is well defined, and confirms the existence of the compound Ag_3Al . The curve between this point and the silver end of the series shows a slight change of slope at approximately 96 per cent. Ag, corresponding to the limit of the α solid solution of aluminium in silver. The resistivity-composition curve does not indicate the presence of a phase boundary at 91.5 per cent. Ag.

The variation of temperature coefficient of resistance with composition is given in graph 3, which again shows changes of slope at 85, 88.5, 92.3 and 96.0 Ag. The limit of the α phase at 96.1 per cent. Ag is more clearly shown in this graph than in graph 2. In the $(\text{Al}+\gamma)$ phase and in the region extending from 85 to 88.5 per cent. Ag an increase in resistivity corresponds to a decrease in temperature coefficient, and in the α phase (96 to 100 per cent. Ag) a decrease in resistivity is accompanied by an increase in temperature coefficient. In the region extending from 88.5 to 92.3 per cent. Ag a sharp increase in resistivity is accompanied by a sharp increase in temperature coefficient, and in the region extending from 92.3 to 96.1 per cent. Ag the resistivity and temperature coefficient of resistance decrease together.

The thermoelectric power-composition curve (graph 4) also shows changes of slope at 85, 88.5, 92.3 and 96.1 per cent. Ag, and the discontinuity at the composition (92.3 per cent. Ag) corresponding to the compound Ag_3Al is very well defined. This curve, however, shows an additional discontinuity at 99.5 per cent. Ag, and it is interesting to note that this discontinuity is also obtained in the Hall coefficient curve, which is very similar in shape to the thermoelectric power curve. Since no evidence of a phase change at 99.5 per cent. Ag had been obtained as the result of X-ray and thermal investigations, two of the alloys having compositions of 97.0 and 99.5 per cent. Ag were subjected to a further period of annealing, but no change in the values of the physical constants was produced. Investigations carried out in this laboratory on the physical properties of the Cu-Sn system have shown that the thermoelectric power-composition curve shows a similar peculiarity when compared with the other curves.

The graph (5) showing the relation between Hall coefficient and composition also indicates the presence of the same phase boundaries, and the discontinuity at the composition of the compound Ag_3Al is again very well defined. An examination of the graph shows that alloys containing about 73.5 and 89.5 per cent. Ag have a zero Hall coefficient.

The results of the present investigation with respect to the position of the phase boundaries at 88.5, 92.3 and 96.1 per cent. Ag are in agreement with Petrenko's equilibrium diagram. This investigation, however, indicates the presence of a fourth boundary at a composition of approximately 85 per cent. Ag, but an insufficient number of alloys were cast in this region to fix its position accurately. The X-ray investigations of Westgren and Bradley fix the boundary between the $(\text{Al}+\gamma)$ and

γ phases at 85.6 per cent. Ag. According to these investigators, the β' phase (Ag_3Al) is in equilibrium with the γ phase in the range 92.3 to 91.5 per cent. Ag by weight, but in the present investigation no evidence has been obtained of a discontinuity in any of the curves at the composition 91.5 per cent. Ag.

The variation of the various physical properties with composition indicates that the boundary of the α solid solution is situated at the composition 96 per cent. Ag, and this result is in agreement with the composition given in Petrenko's diagram. The X-ray investigations of Westgren and Bradley and of Phelps and Davey⁽¹⁴⁾ indicate that the boundary is situated at 94.1 and 94.6 per cent. Ag respectively. Ageew and Shoyket⁽¹⁵⁾ applied the X-ray method to determine the boundary of the α phase, and found that it corresponds to a composition of 6.52 per cent. by weight of Al at 500° C., and to 5.10 per cent. by weight of Al at 200° C. This diminution in solubility may, however, not continue to room temperature.

All the curves show the existence of sharp peaks or troughs corresponding to the compound Ag_3Al , and also changes of slope, which are not so definite, corresponding to Ag_3Al_2 and Ag_2Al . According to the work of Westgren and Bradley, Ag_3Al_2 is situated on one of the boundaries of the homogeneous γ phase, and Ag_2Al is situated in the homogeneous γ phase. This homogeneous γ phase extends over the range 85.65 to 91.5 per cent. by weight of silver, and is a solid solution of close-packed hexagonal structure.

References.

- (1) Petrenko, *Zeits. Anorg. Chem.* (46) i. p. 49 (1905).
- (2) Westgren and Bradley, *Phil. Mag.* vi. p. 280 (August, 1928).
- (3) Phelps and Davey, *Am. Inst. Mining and Met. Eng. Tech. Publ.* no. 443 (1932).
- (4) Hoar and Rowntree, *Journ. Inst. Metals*, xlv. p. 119 (1931).
- (5) Ageew and Shoyket, *Journ. Inst. Metals*, lii. p. 119 (1933).
- (6) Crepaz, *Atti III. Congr. naz. Chim. pura appl. Firenze e Toscana*, p. 371 (1929); *Chem. Zentr.* cii. i. p. 2158 (1931).
- (7) Le Grix and Broniewski, *Rev. Met.* x. p. 105 (1913).
- (8) Hanson and Slater, *Journ. Inst. Metals*, lvi. p. 103 (1935).
- (9) Thomas and Evans, *Phil. Mag.* xvi. p. 329 (1933).
- (10) Westgren and Bradley, *loc. cit.*
- (11) Crepaz, *loc. cit.*
- (12) Le Grix and Broniewski, *loc. cit.*
- (13) 'A Treatise on Inorganic and Theoretical Chemistry,' vol. v. pp. 232, 233.
- (14) Phelps and Davey, *loc. cit.*
- (15) Ageew and Shoyket, *loc. cit.*

Physics Department,
University College of Swansea.
December, 1939.

XX. *The Approximate Theory of the Displacements produced in Encastré Beams due to the Elasticity of the Supports.*

By S. T. NEWING, B.Sc., A.R.C.S., D.I.C.,
The University, Bristol *.

[Received September 25, 1942.]

1. *Introduction.*

THE usual theory of clamped and encastré beams takes no account of the elastic properties of the medium, in which one end of the beam is fixed. It is customary to describe the method of fixing that end by boundary conditions which express the displacement y , and the slope $\frac{dy}{dx}$ of the beam, at the fixed end $x=0$. These values are generally taken to be zero. On the usual assumptions with regard to the beam †, a solution of the differential equation

$$\frac{d^2}{dx^2} \left(EI \frac{d^2 y}{dx^2} \right) = p(x) \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.1)$$

is found, which conforms with these conditions.

In the above equation, E , I , y and x have their usual significance, and $p(x)$ is the loading per unit length of the beam.

The object of this investigation is to estimate the displacement produced in a cantilever or a doubly-encastré beam if, instead of assuming the conditions stated above, the foundations into which the fixed end of the beam is built are assumed to be elastic, and to abut on rigid supports as shown in fig. 1.

No account is taken of the restraining forces which would act within the length of the foundation over the vertical faces of the beam. It is considered unlikely that these would be comparable in magnitude with those acting over the horizontal faces, and the problem has therefore been simplified by assuming that the width of the elastic supporting columns is equal to that of the beam.

The displacement of the neutral axis of the beam is given by equation (1.1), but $p(x)$ will have one mathematical form for the portion OB within the foundations, and another, dependent on the applied loading, between B and C. There must be continuity of displacement, slope, bending moment, and shearing force at B to relate the solutions of (1.1) for the two portions of the beam. Within the portion OB it is

* Communicated by the Author.

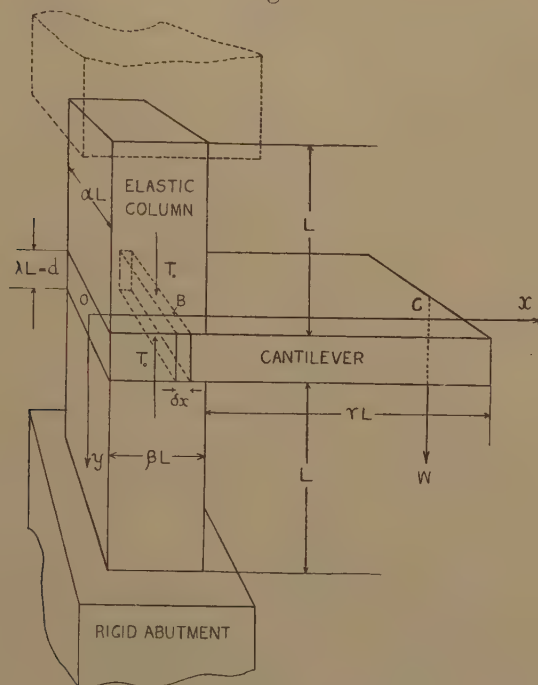
† For a statement of these assumptions, see 'Analysis of Structures,' by Baker and Pippard, p. 22, § 3.3.

assumed that the effect of the foundation columns can be represented by supposing that the beam is constrained by a large number of springs. These are supposed to be attached at one end to the rigid abutments, and at other to the horizontal surfaces of the beam. The actions of these springs can be represented by a distributed force equal to Ky per unit length, where K is the effective spring constant. In the absence of any loading on the beam within the portion OB, the differential equation (1.1) takes the form

$$EI \frac{d^4 y}{dx^4} + Ky = 0, \quad \dots \dots \dots (1.2)$$

provided that the beam is of uniform cross-section and of negligible weight.

Fig. 1.



Method of supporting cantilever.

If the beam is a heavy one, then the displacement is determined from the equation

$$EI \frac{d^4 y}{dx^4} + Ky = W(x), \quad \dots \dots \dots (1.3)$$

where $W(x)$ is the weight per unit length of the beam at distance x from O .

For the portion BC, equations (1.2) and (1.3) are replaced by

$$EI \frac{d^4 y}{dx^4} = 0 \quad \dots \dots \dots (1.4)$$

for a light beam loaded at C, and for a heavy beam by

$$EI \frac{d^4 y}{dx^4} = W(x), \quad \dots \quad (1.5)$$

where $W(x)$ is now the total loading per unit length, *i. e.* applied loading plus the weight.

The displacement of the beam within the region OB has been determined for a light beam with an isolated load at C, and also for a heavy beam similarly loaded. In order to increase the effective support given to the beam it is assumed that, initially, there is a lack of fit between the beam and the elastic foundation, so that both are in compression in the y direction. Clearly it is possible to relieve the forced fit on the upper horizontal surface at B by applying a sufficiently large load at C. In fact, the forced fit may be relieved along a certain distance to the left of B, and in this connexion the dependence of the applied load upon the length of the beam which has been "freed" on its upper surface has been investigated.

The additional displacement at the end of the beam due to the yielding of the elastic foundations has been calculated for cantilevers, for beams which project on both sides of the foundations, for doubly encastred beams under various conditions of loading, and for foundations of various sizes. Simple approximate formulæ have been derived for use in connexion with beams and foundations, whose dimensions comply with the important inequality (4.4).

2. Stiffness of the Elastic Foundations.

The height of the foundation columns is denoted by L . The vertical rectangular cross-section of the beam has height $d = \lambda L$ and width αL . The foundation columns have a rectangular horizontal cross-section αL by βL . The length of the portion of the beam which projects beyond the foundations is γL , its total length being $(\gamma + \beta)L$.

The original height of the aperture between the columns is denoted by " t ," so that $(d - t)$ is the original lack of fit. T_0 denotes the normal force per unit area, assumed to be uniform, over the horizontal surfaces of an element δx of the beam distant x from O when the beam has been forced into the foundations. If y is the displacement of the beam at distance x from O, then the downward force on the upper surface of the element is

$$(T_0 - E_1 y/L) \cdot \delta x \cdot d.$$

In this expression E_1 denotes the stiffness of the springs which replace the elastic foundations, and is taken to be the same as Young's Modulus for the material of which the foundations are composed. Similarly, the upward force on the lower face of the element is

$$\left(T_0 + \frac{E_1 y}{L}\right) \cdot \delta x \cdot d.$$

The resultant upward force per unit length is, therefore,

$$2E_1 y d/L = -p(x).$$

(4) At $x=L$ the shearing force is W , so that $\frac{d^3y}{dx^3} = -W/EI$.

The first three derivatives of the displacement function y , derived from (2.2), are given by

$$\frac{1}{\mu} \frac{dy}{dx} = \exp(\mu x)[(A+B) \cos \mu x + (B-A) \sin \mu x] \\ + \exp(-\mu x)[(D-C) \cos \mu x - (C+D) \sin \mu x],$$

$$\frac{1}{2\mu^2} \frac{d^2y}{dx^2} = \exp(\mu x)[B \cos \mu x - A \sin \mu x] - \exp(-\mu x)[D \cos \mu x - C \sin \mu x],$$

and

$$\frac{1}{2\mu^3} \frac{d^3y}{dx^3} = \exp(\mu x)[(B-A) \cos \mu x - (B+A) \sin \mu x] \\ + \exp(-\mu x)[(D+C) \cos \mu x + (D-C) \sin \mu x].$$

Condition (1) above gives $B-D=0$, $\therefore B=D$.

Condition (2) gives $B-A+D-C=0$, $\therefore A=2B+C$.

The constants B and C are determined from the remaining boundary conditions (3) and (4), which yield the equations

$$3WL/2\mu^2 EI = \exp(\mu L)[B \cos \mu L - (2B+C) \sin \mu L] \\ - \exp(-\mu L)[B \cos \mu L - C \sin \mu L], \quad (3.1)$$

and

$$W/2\mu^3 EI = \exp(\mu L)[(B+C) \cos \mu L + (3B+C) \sin \mu L] \\ - \exp(-\mu L)[(B+C) \cos \mu L + (B-C) \sin \mu L]. \quad (3.2)$$

Now $\mu L = 8.801116$, so that

$$\exp(\mu L) \simeq 6.64 \times 10^3, \quad \text{whereas} \quad \exp(-\mu L) \simeq 1.51 \times 10^{-4},$$

thus in evaluating the constants B and C it is hardly necessary to include terms which involve the negligibly small negative exponentials. Since different boundary conditions at $x=L$ will arise in other problems to be considered, the above equations (3.1) and (3.2) are solved in terms of the bending moment M and shearing force W at $x=L$.

These equations may be rearranged as

$$M \exp(-\mu L)/2\mu^2 EI = B(\cos \mu L - 2 \sin \mu L) - C \sin \mu L,$$

$$\text{and } W \exp(-\mu L)/2\mu^3 EI = B(\cos \mu L + 3 \sin \mu L) + C(\sin \mu L + \cos \mu L),$$

in which the negative exponentials have been neglected in comparison with the positive ones.

If the left-hand sides of these equations are denoted respectively by p_1 and p_2 , it is easily verified that

$$B = p_1(\sin \mu L + \cos \mu L) + p_2 \sin \mu L,$$

$$\text{and } -C = p_1(3 \sin \mu L + \cos \mu L) + p_2(2 \sin \mu L - \cos \mu L),$$

equations which enable B and C to be evaluated in terms of M , W , E and I for $\mu L = 8.801116$.

It is, however, possible to calculate the slope and displacement of the beam at $x=L$, *without* evaluating B and C , provided the value of μL

allows the negative exponential, $\exp(-\mu L)$, to be neglected. For if w and w_1 denote respectively the displacement and slope at $x=L$, then

$$\begin{aligned} w &= \exp(\mu L)[A \cos \mu L + B \sin \mu L] \\ &= \exp(\mu L)[(2B+C) \cos \mu L + B \sin \mu L]. \end{aligned}$$

Now $2B+C=p_1(\cos \mu L - \sin \mu L) + p_2 \cos \mu L$,

so that

$$(2B+C) \cos \mu L = p_1[\cos^2 \mu L - \sin \mu L \cos \mu L] + p_2 \cos^2 \mu L,$$

and

$$B \sin \mu L = p_1[\sin \mu L \cos \mu L + \sin^2 \mu L] + p_2 \sin^2 \mu L.$$

Thus

$$\begin{aligned} w &= \exp(\mu L) \cdot (p_1 + p_2) \\ &= (W + \mu M)/2\mu^3 EI. \end{aligned} \quad (3.3)$$

Similarly, for the displacement w_1 , it may be shown that

$$w_1 = (W + 2\mu M)/2\mu^2 EI. \quad (3.4)$$

The numerical values of the constants A, B, C, D for the dimensions assumed in this section are :—

$$A = -4.159m,$$

$$B = -0.5995m,$$

$$C = -2.960m,$$

$$D = -0.5995m,$$

where $m = WL^3/EI \times 10^6$, and is dimensionally a length.

On the basis of these values, the displacement of the beam within the range $0 \leq x \leq L$ has been calculated, together with the shearing force and bending moment. These numerical values are shown in Table I. and are represented graphically in fig. 3. K_1, K_2, K_3 are constants whose values are shown at the top of the columns in Table I.

It must be emphasized that the equations (3.3) and (3.4) take those forms only if the beam is long enough. They do not depend upon the extent of the original lack of fit, and, indeed, they retain the same form if there is no forced fit initially.

It is clear from examination of the graphs in fig. 3 that the displacement is only appreciable within the second half of the foundation, and considerable only within the fourth quarter, *i. e.* between $x=0.75L$ and $x=L$. The same remark applies to the bending moment and shearing force graphs, showing that the most significant portion of the force system which equilibrates the load, is situated within this fourth quarter of the foundation. For a given set of values of L, E, I and x , the displacement y is proportional to W . It might be supposed that, by a sufficient increase in W , it would be possible to obtain a displacement sufficiently large to relieve the forced fit between the upper surface of the beam and the elastic foundation for any x value, provided only that the stresses set up in the beam were not large enough to cause a fracture of its material. The displacement curve in fig. 3 shows that this supposition is not correct. Between P and B it is possible, subject to the above proviso, to obtain any displacement by making W large

enough, and, therefore, to relieve the forced fit on the upper surface. It is possible that this value of W would lead to such large negative (upward) displacements in the region between Q and P that in some sub-division of that region the forced fit on the lower surface of the beam might be

TABLE I.

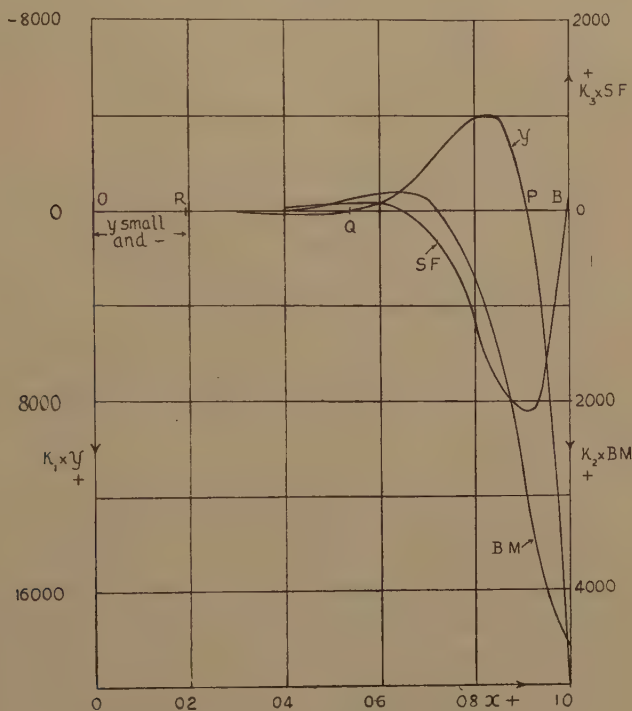
x/L .	$\frac{yEI \times 10^6}{WL^3}$.	$\frac{B. M. \times 10^4}{WL \times 5.937}$.	$\frac{S. F. \times 10^3}{W \times 8.100}$.
0	— 7.119	0	0
0.1	— 8.479	+ 1.432	— 2.377
0.2	+ 1.124	+ 5.682	— 4.003
0.3	+ 47.22	+ 8.386	+ 1.915
0.4	+ 138.1	— 7.888	+ 28.80
0.42	+ 155.3	— 16.14	+ 37.54
0.44	+ 168.1	— 26.64	+ 47.12
0.46	+ 174.0	— 39.65	+ 57.31
0.48	+ 169.5	— 55.21	+ 67.56
0.50	+ 150.5	— 73.16	+ 77.05
0.52	+ 113.6	— 94.35	+ 85.82
0.54	+ 50.05	— 115.3	+ 89.92
0.56	— 42.31	— 137.8	+ 90.31
0.58	— 170.6	— 159.7	+ 84.22
0.60	— 340.3	— 178.9	+ 69.27
0.62	— 562.2	— 195.2	+ 43.41
0.64	— 823.2	— 199.2	+ 2.274
0.66	— 1142	— 192.9	— 55.66
0.68	— 1509	— 169.7	— 133.9
0.70	— 1922	— 124.4	— 235.4
0.72	— 2364	— 50.53	— 360.8
0.74	— 2820	+ 58.13	— 515.7
0.76	— 3261	+ 208.4	— 696.2
0.78	— 3642	+ 406.3	— 900.2
0.80	— 3922	+ 658.8	— 1126
0.82	— 4027	+ 967.8	— 1362
0.84	— 4000	+ 1337	— 1597
0.86	— 3381	+ 1763	— 1815
0.88	— 2421	+ 2236	— 1989
0.90	— 877.8	+ 2747	— 2094
0.92	+ 1375	+ 3269	— 2080
0.94	+ 4484	+ 3769	— 1911
0.96	+ 8568	+ 4203	— 1530
0.98	+ 13745	+ 4506	— 872.7
1.00	+ 20102	+ 4607	+ 125.6

relieved also. In view of the relative values of the displacements in the sections QP and PB of the beam, it is unlikely that the forced fit anywhere on the lower surface between Q and P will be relieved unless the point at which the forced fit on the upper surface has been relieved lies very close to P .

The same argument may be extended to include the other portions of OP. Thus it might be possible to obtain relief of the forced fit on the lower surface within some regions of OR and QP, and on the upper surface for some regions within RQ and PB. However large W is made, apart from values involving the fracture of the material, it is not possible to relieve the forced fits on either surface at the points R, Q, P where the displacement is zero for any value of W .

The values of W which will cause a given displacement for various x values between P and B have been calculated in section 6, and by

Fig. 3.



Displacement, B. M. and S. F. diagrams.

equating this displacement to the original lack of fit, the values of W are found which will relieve the forced fit for points between P and B.

4. Foundation Columns of Various Sizes.

Suppose that, with the other dimensions of the foundations unchanged, the length of the columns is taken to be βL (see fig. 1) instead of L , as in the previous section. As before, the bending moment and shearing force at B, now $x = \beta L$, are denoted by M and W respectively.

The solution of equation (2.1) has the form given by (2.2), in which the constant μ is $8.801116/L$, as before. The constants appearing in

the displacement function are to be determined by the following conditions :—

$$\text{at } x=0, \quad (1) \quad \frac{d^2y}{dx^2}=0; \quad (2) \quad \frac{d^3y}{dx^3}=0;$$

$$\text{at } x=\beta L, \quad (3) \quad \frac{d^2y}{dx^2}=M/EI; \quad (4) \quad \frac{d^3y}{dx^3}=-W/EI.$$

The conditions at $x=0$ give, as before, $A=2B+C$, $D=B$; while the equations to determine B and C become

$$M \exp(-\mu\beta L)/2\mu^2 EI = B \cos \mu\beta L - (2B+C) \sin \mu\beta L,$$

$$\text{and} \quad W \exp(-\mu\beta L)/2\mu^3 EI = (B+C) \cos \mu\beta L - (3B+C) \sin \mu\beta L,$$

provided that $\exp(-\mu\beta L)$ can be neglected in comparison with $\exp(\mu\beta L)$. In that case, the values of the constants are easily determined, and the displacement w and slope w_1 at $x=\beta L$ are given by

$$w = (W + \mu M)/2\mu^3 EI, \quad \dots \dots \dots (4.1)$$

$$w_1 = (W + 2\mu M)/2\mu^2 EI. \quad \dots \dots \dots (4.2)$$

Comparison of these results with (3.3) and (3.4) for the previous case, in which $\beta=1$, makes it clear that the same displacement and slope are obtained at the extremity of the foundation, no matter what its length (βL) may be, provided that it is large enough for $\exp(-\mu\beta L)$ to be neglected in comparison with $\exp(\mu\beta L)$. Now the values of w and w_1 are a measure of the extent to which the beam is fixed by the foundation columns. No extra support is afforded then to the beam by using foundations which are much in excess of that length which allows the negative exponential to be neglected. For example, if it is decided to neglect $\exp(-\mu\beta L)$ when it is less than $\exp(\mu\beta L)/100$, β must be such as to make $\mu\beta L$ either greater than or equal to about 2.303. Since $\mu L=8.801116$, the required values of β must be greater than about 0.26. In that case nothing is to be gained in the way of additional support for the beam by making the length of the foundation columns more than about a quarter of their height. In this connexion it is instructive also to calculate the displacement at the end of the beam or cantilever, due merely to the elastic nature of the foundations. The quantity γ is taken to be 3, so that the length of the beam projecting beyond the foundations is $3L$.

The displacement δ at the end of the cantilever due to the yielding of the foundation is given by

$$\delta = w + 3w_1 L,$$

in the case of a light beam. The values of δ , w , and $w_1 L$ shown in Table II. have been calculated for various values of β , without making the approximation neglecting $\exp(-\mu\beta L)$. For purposes of comparison, it should be noted that the displacement at the end of a light cantilever of length $3L$, carrying a load W at its free end, and fixed rigidly so that w and w_1 are both zero, is $9WL^3/EI$. At best, therefore, the foundation is bound to increase the ideal displacement by about one-ninth of the

usual value, since the limiting value of δ for a very long beam is slightly in excess of WL^3/EI . It should be noted that the foundation has been assumed to have the same modulus E as the beam. Thus, if the beam is constructed of steel, the value of E_1 is certainly less than E_2 , unless the foundations also are of steel. Consequently for foundations of most other materials the fraction one-ninth, quoted above, will be considerably increased.

The effect of altering the height of the foundation columns will now be considered. This quantity will now be denoted by ϵ^4L . The differential equation (2.1) now becomes

$$\frac{d^4y}{dx^4} + \frac{2d}{1L\epsilon^4} \cdot y = 0,$$

with $E_1 = E_2$, as before.

TABLE II.

β	wEI/WL^3	w_1EI/WL^3	$EI\delta/WL^3$
0.1	0.07827	1.6371	4.9897
0.2	0.02769	0.5274	1.6098
0.25	0.02148	0.3597	1.1005
0.3	0.02031	0.3497	1.0695
0.5	0.02016	0.3480	1.0646
0.6	0.02010	0.3473	1.0620
0.8	0.02010	0.3473	1.0620
1.0	0.02010	0.3473	1.0620
∞	0.02010	0.3473	1.0620

The solution of this equation is given by (2.2) as before, except that $\mu = 8.801116/\epsilon L$. If the further assumptions are made that $\gamma = 3$, and that the value of β is such as to allow the negative exponential terms to be neglected, then we may write for the displacement w and slope w_1 at $x = \beta L$ the relations

$$w = W(3\mu L + 1)/2\mu^3 EI,$$

and

$$w_1 = W(6\mu L + 1)/2\mu^2 EI.$$

On substituting for μ , these become

$$w = WL^3(0.01937\epsilon^2 + 0.0007337\epsilon^3)/EI,$$

$$w_1 = WL^2(0.3409\epsilon + 0.006455\epsilon^2)/EI.$$

Thus δ , the displacement at the free end due to the yielding of the elastic foundation columns, is given by

$$\delta = w + 3w_1L,$$

$$= WL^3(1.0227\epsilon + 0.03874\epsilon^2 + 0.0007337\epsilon^3)/EI. \quad \dots \quad (4.3)$$

This shows that, for short foundation columns, *i. e.* for small values of ϵ , the displacement at the free end is approximately proportional to ϵ , *i. e.* to the one-fourth power of the ratio of the height of the foundation

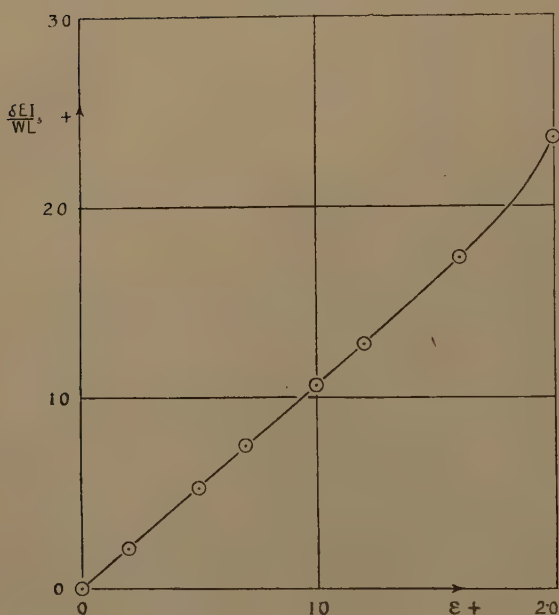
columns to their length. The values of δ for various values of ϵ have been calculated from (4.3) and are illustrated in Table III. and fig. 4.

If the vertical cross-section of the beam retains its square shape, the only other possible variation in the problem, apart from the variation of γ , and a distributed load between B and C, both of which are trivial, is to vary the size of the square section.

TABLE III.

ϵ .	$\delta EI/WL^3$.
0	0
0.2	0.2061
0.5	0.5261
0.7	0.7447
1.0	1.0623
1.2	1.2845
1.6	1.7388
2.0	2.3619

Fig. 4.



Effect of height of columns on yielding of foundations.

Thus a beam is now considered in which $\alpha = \lambda$, and the height and length of the foundation columns are taken as $\epsilon^4 L$ and βL respectively. Equation (2.1) takes the form

$$\frac{d^4 y}{dx^4} + \frac{24}{\alpha^3 \epsilon^4 L^4} \cdot y = 0;$$

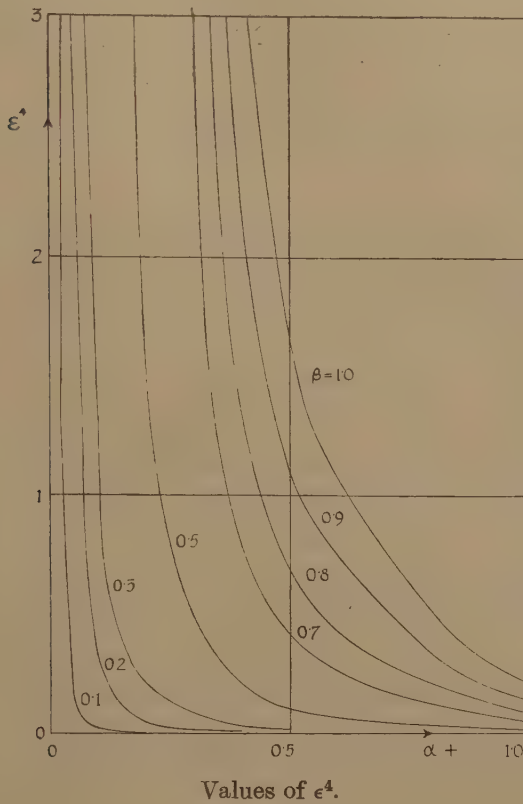
since $I = \alpha^4 L^4/12$ and $d = \alpha L$.

TABLE IV.—Values of ϵ^4 .

$\frac{\beta}{\alpha}$	0.1	0.2	0.3	0.5	0.7	0.8	0.9	1.0
0.025	1.3617	21.780	110.30	851.06	3269.4	5577.5	8934.1	13617
0.050	0.1702	2.7234	13.787	106.38	408.68	697.19	1116.8	1702.1
0.100	0.0213	0.3404	1.7234	13.298	51.085	87.149	139.60	212.77
0.200	0.0026	0.0425	0.2154	1.6622	6.3856	10.089	17.495	26.596
0.400	0.0003	0.0053	0.0269	0.2078	0.7982	1.3617	2.1812	3.3245
0.500	0.0002	0.0027	0.0138	0.1064	0.4087	0.6972	1.1168	1.7021
0.800	—	0.0008	0.0042	0.0325	0.1247	0.2128	0.3408	0.5194
1.000	—	0.0003	0.0017	0.0133	0.0511	0.0871	0.1396	0.2128

The values of ϵ^4 , which seem to be of reasonable practical interest are those lying between 0.01 and 2, and these are enclosed within the thick line in Table IV. The same values are also illustrated in fig. 5.

Fig. 5.



5. Displacement of a Heavy Beam.

The beam, which in sections 3 and 4 was regarded as weightless, is now considered to have a uniform weight w_0 per unit length; β and ϵ are taken to be unity and α to be 0.1 as in section 3, so that $\mu L = 8.801116$.

The equation for the displacement is

$$\frac{d^4 y}{dx^4} + 4\mu^4 y = w_0/EI,$$

of which the solution is

$$y = \exp(\mu x)[A \cos \mu x + B \sin \mu x] + \exp(-\mu x)[C \cos \mu x + D \sin \mu x] + w_0/4\mu^4 EI.$$

Now for the assumed values of the dimensions of the beam and foundations,

$$4\mu^4 EI = E/5,$$

so that

$$y = \exp(\mu x)[A \cos \mu x + B \sin \mu x] \\ + \exp(-\mu x)[C \cos \mu x + D \sin \mu x] + 5w_0/E.$$

The addition of the constant term to the solution does not affect the evaluation of the constants A, B, C and D, in conformity with the same boundary conditions as in section 3. since these conditions involve derivatives of y to which the constant term makes no contribution.

The displacement at the free end of the beam due to the yielding of the foundation columns takes the same form as before, with the addition of the constant term $5w_0/E$.

Thus

$$\delta = 5w_0/E + [W + \mu M]/2\mu^3 EI + 3L[W + 2\mu M]/2\mu^3 EI,$$

where M and W are the bending moment and shearing force at $x=L$. If W_1 is the total weight of the beam,

$$W_1 = w_0(1+3)L = 4w_0L,$$

and

$$W = 3w_0L = \frac{3}{4}W_1,$$

so that $\delta = 5W_1/4EL + [\mu M + \frac{3}{4}W_1]/2\mu^3 EI + 3L[2\mu M + \frac{3}{4}W_1]/2\mu^2 EI$.

6. Relief of the Forced Fit on the Upper Surface of a Light Beam between B and P.

If the displacement of the beam at some point D for which $x=p\beta L$, is equal to half the original lack of fit between the beam and its foundations, then the upper surface of the beam will just be relieved from the forced fit at this point. The shearing force and bending moments at D are denoted by W_1 and M_1 , respectively, their values at P by W and M, as before, where W is the weight attached to the free end of the beam.

An examination of fig. 3 shows that OD must exceed about 0.9L. In consequence both $\exp(-\mu p\beta L)$ and $\exp(-\mu\beta L)$ may be neglected in comparison with their reciprocals. It is required to find the value of p , for a given value of W, supposing, of course, that this value of W lies within the range of values for which the forced fit on the upper surface will be relieved somewhere between P and B. Although this cannot be done directly, it is possible to find W for certain assumed values of p (which exceed 0.908646, the x value for P), by the method described below, so that a graph may be constructed connecting p and W. From this the value of p for a given W may be found.

Since $\exp(-\mu p\beta L)$ may be neglected, it is possible to use the approximate formulæ for the displacement and slope at D. Thus at D,

$$y = (\mu M_1 + W_1)/2\mu^3 EI, \\ \frac{dy}{dx} = (2\mu M_1 + W_1)/2\mu^2 EI, \\ \frac{d^2y}{dx^2} = M_1/EI,$$

$$\frac{d^3y}{dx^3} = -W_1/EI,$$

where $\mu L = (6/\alpha^3 \epsilon^4)^{\frac{1}{2}}$, as in section 4.

The general expression for y is given by (2.2) with the above value of μ .

For the portion DB of the beam, displacements are resisted by one of the foundation columns only, so that the differential equation (2.1) becomes

$$\frac{d^4y}{dx^4} + \frac{d}{L} \cdot y = 0.$$

If $\nu = \mu(2)^{-\frac{1}{2}}$, then since $\mu L = 8.801116$, $\nu L = 7.400828$, and the solution for y is

$$y = \exp(\nu x)[A' \cos \nu x + B' \sin \nu x] + \exp(-\nu x)[C' \cos \nu x + D' \sin \nu x]$$

In this solution, while it is clear that at $x = p\beta L$, the negative exponential $\exp(-\nu p\beta L)$ is small compared with $\exp(\nu p\beta L)$, there are no longer any such boundary conditions as in section 3, which imply that the four constants have the same order of magnitude, and that in consequence the terms involving the negative exponentials may be neglected. In fact, all four constants must be retained in order that the four conditions of continuity at D, viz. for displacement, slope, bending moment and shearing force, may be satisfied, and so determine the values of the constants.

The forms for $\frac{dy}{dx}$, $\frac{d^2y}{dx^2}$, $\frac{d^3y}{dx^3}$ may be written down from those in section 3 by substituting ν for μ and using primed constants. Writing $\phi = \nu p\beta L$, the continuity conditions are as follows:

$$(\mu M_1 + W_1)/2\mu^3 EI = \exp(\phi)[A' \cos \phi + B' \sin \phi] \\ + \exp(-\phi)[C' \cos \phi + D' \sin \phi],$$

$$(2\mu M_1 + W_1)/2\nu\mu^2 EI = \exp(\phi)[(A' + B') \cos \phi + (B' - A') \sin \phi] \\ + \exp(-\phi)[(D' - C') \cos \phi - (C' + D') \sin \phi],$$

$$M_1/2\nu^2 EI = \exp(\phi)[B' \cos \phi - A' \sin \phi] \\ - \exp(-\phi)[D' \cos \phi - C' \sin \phi],$$

$$-W_1/2\nu^3 EI = \exp(\phi)[(B' - A') \cos \phi - (B' + A') \sin \phi] \\ + \exp(-\phi)[(C' + D') \cos \phi + (D' - C') \sin \phi].$$

For given values of p and β , ϕ is a definite number, and these four equations suffice to determine the constants A' , B' , C' and D' in terms of the unknown, assumed, bending moment and shearing force M_1 and W_1 . These two quantities are simply related to M and W , and may be evaluated as follows. From the conditions of equilibrium of the portion DC of the beam, it is clear that

$$W_1 + \int_{p\beta L}^{\beta L} \frac{E y d}{L} dx = W, \quad \dots \dots \dots (6.1)$$

and $M_1 + \int_{p\beta L}^{\beta L} \frac{E y d}{L} (x - p\beta L) dx = M = 3WL. \dots \dots \dots (6.2)$

Since the integrals involve the assumed form for y , in which the constants A' , B' , C' , D' have been evaluated already in terms of M_1 and W_1 , it follows that the above two equations reduce to a pair of simultaneous algebraic equations to determine M_1 and W_1 in terms of M and W , after the necessary integration has been performed. The integration is simple but tedious, and involves the following standard results, in which constants of integration have been omitted :

$$\begin{aligned}\int \exp(\nu x) \cos \nu x \, dx &= \exp(\nu x) [\cos \nu x + \sin \nu x] / 2\nu = f_1(x), \\ \int \exp(\nu x) \sin \nu x \, dx &= \exp(\nu x) [\sin \nu x - \cos \nu x] / 2\nu = f_2(x), \\ \int \exp(-\nu x) \cos \nu x \, dx &= \exp(-\nu x) [\sin \nu x - \cos \nu x] / 2\nu = f_3(x), \\ \int \exp(-\nu x) \sin \nu x \, dx &= \exp(-\nu x) [\sin \nu x + \cos \nu x] / 2\nu = f_4(x), \\ \int x \exp(\nu x) \cos \nu x \, dx &= \exp(\nu x) [\nu x (\cos \nu x + \sin \nu x) - \sin \nu x] / 2\nu^2 = g_1(x), \\ \int x \exp(\nu x) \sin \nu x \, dx &= \exp(\nu x) [\nu x (\sin \nu x - \cos \nu x) + \cos \nu x] / 2\nu^2 = g_2(x), \\ \int x \exp(-\nu x) \cos \nu x \, dx &= \exp(-\nu x) [\nu x (\sin \nu x - \cos \nu x) + \sin \nu x] / 2\nu^2 = g_3(x), \\ \int x \exp(-\nu x) \sin \nu x \, dx &= -\exp(-\nu x) [\nu x (\sin \nu x + \cos \nu x) + \cos \nu x] / 2\nu^2 = g_4(x).\end{aligned}$$

Now $Ed/L = EI \cdot \frac{d}{IL} = 4EI\nu^4$, so that the equations (6.1), (6.2) become

$$\begin{aligned}W_1 + 4EI\nu^4 \int_{p\beta L}^{\beta L} y \, dx &= W, \\ M_1 + 4EI\nu^4 \int_{p\beta L}^{\beta L} y(x - p\beta L) \, dx &= 3WL.\end{aligned}$$

Writing $\nu\beta L = \theta$, these equations take the final form

$$\begin{aligned}W = W_1 + 4EI\nu^4 [A' \{f_1(\theta) - f_1(\phi)\} + B' \{f_2(\theta) - f_2(\phi)\} \\ + C' \{f_3(\theta) - f_3(\phi)\} + D' \{f_4(\theta) - f_4(\phi)\}],\end{aligned}$$

and

$$\begin{aligned}3WL = M_1 + 4EI\nu^4 [A' \{g_1(\theta) - g_1(\phi)\} + B' \{g_2(\theta) - g_2(\phi)\} \\ + C' \{g_3(\theta) - g_3(\phi)\} + D' \{g_4(\theta) - g_4(\phi)\}],\end{aligned}$$

which determine M_1 and W_1 when the values of A' , B' , C' , D' , in terms of these quantities, have been inserted in the right-hand sides. Then by substituting these values in the equation

$$y = (\mu M_1 + W_1) / 2\mu^3 EI,$$

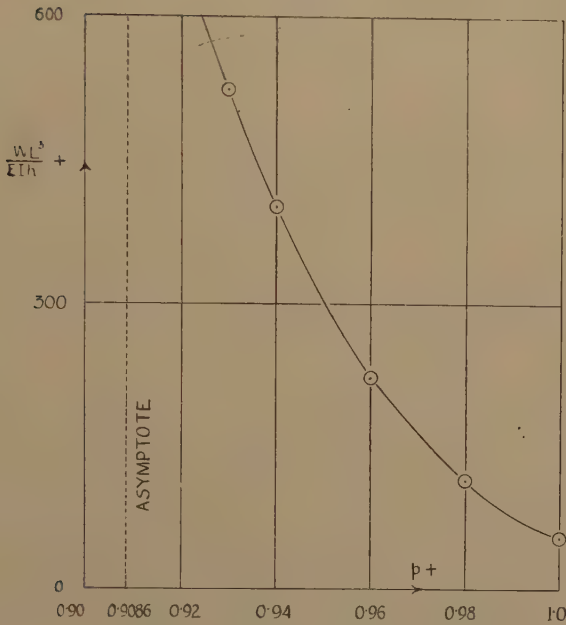
the displacement at $x = p\beta L$ may be calculated in terms of W . This laborious series of calculations was performed for the following values of p ; 1, 0.98, 0.96, 0.94, 0.93, and the results are shown in Table V. and fig. 6. If " h " is used to denote half the original lack of fit between the beam and its foundations, W the load required at the end of the beam to produce this displacement at $x = p\beta L$, then Table V. and fig. 6 show the relation between p and WL^3/EIh .

The last value in Table V. corresponds with the point P in fig. 3. If there is no original lack of fit, then fig. 3 shows that the beam will be held at its upper surface between Q and P, and also between O and R, and at its lower surface between R and Q and also between P and B.

TABLE V.

p .	WL^3/EIh .
1.00	54
0.98	114
0.96	222
0.94	401
0.93	523
0.908646 . . .	∞

Fig. 6.



Variation of freed length with "W."

7. The Displacement of a Light Beam whose Central Portion is held between Two Elastic Foundations.

It is now supposed that the light beam considered in previous sections projects to equal distances $3L$ on both sides of the foundation columns, whose length is βL , height $\epsilon^4 L$, and width αL , as before. The vertical cross-section of the beam is taken to be a square of side αL .

At the free ends of the beam there are applied equal loads W . Within

the foundation, equation (2.1) determines the displacement y of the beam. In this section the origin is taken half-way along the foundation, whose extremities have x values $\pm\beta L/2$. As in section 3, the displacement y is given by

$$y = \exp(\mu x)[A \cos \mu x + B \sin \mu x] + \exp(-\mu x)[C \cos \mu x + D \sin \mu x],$$

where $\mu L = [6/\alpha^3 \epsilon^4]^{\frac{1}{2}}$.

The symmetry of the problem requires $y(x) = y(-x)$ and, in addition, the boundary conditions at

$$x = \pm \frac{\beta L}{2}; \text{ viz. } \frac{d^2 y}{dx^2} = M/EI \text{ and } \frac{d^3 y}{dx^3} = -W/EI,$$

must be satisfied, M and W having their usual significance.

The symmetry requires that $A = C$ and $D = -B$, and to determine B and C the boundary conditions give the equations

$$3WL/2\mu^2 EI = \exp(\mu\beta L/2) \left[B \cos \frac{\mu\beta L}{2} - C \sin \frac{\mu\beta L}{2} \right] \\ + \exp(-\mu\beta L/2) \left[C \sin \frac{\mu\beta L}{2} - B \cos \frac{\mu\beta L}{2} \right],$$

and

$$W/2\mu^3 EI = \exp(\mu\beta L/2) \left[(C - B) \cos \frac{\mu\beta L}{2} + (B + C) \sin \frac{\mu\beta L}{2} \right] \\ + \exp(-\mu\beta L/2) \left[(B - C) \cos \frac{\mu\beta L}{2} + (B + C) \sin \frac{\mu\beta L}{2} \right].$$

If β is taken as unity, then $\exp(-\mu L/2)$ may be neglected in comparison with its reciprocal, provided that

$$\frac{1}{2}[6/\alpha^3 \epsilon^4]^{\frac{1}{2}} > 2.30.$$

If this inequality is satisfied, then

$$3WL/2\mu^2 EI = \exp(\mu L/2) \left[B \cos \frac{\mu L}{2} - C \sin \frac{\mu L}{2} \right] \quad . \quad . \quad . \quad (7.1)$$

$$\text{and} \quad W/2\mu^3 EI = \exp(\mu L/2) \left[(C - B) \cos \frac{\mu L}{2} + (B + C) \sin \frac{\mu L}{2} \right]. \quad (7.2)$$

By addition of these equations,

$$\exp(\mu L/2) \left[C \cos \frac{\mu L}{2} + B \sin \frac{\mu L}{2} \right] = W(3\mu L + 1)/2\mu^3 EI,$$

and this is the displacement at $x = L/2$.

If (7.1) is multiplied by 2 and added to (7.2), and the resulting sum multiplied by μ , the quantity

$$\mu[3WL/\mu^2 EI + W/2\mu^3 EI] = \mu \exp\left(\frac{\mu L}{2}\right) \left[(C + B) \cos \frac{\mu L}{2} + (B - C) \sin \frac{\mu L}{2} \right]$$

is obtained, which is the slope at $x = L/2$.

Denoting this displacement and slope by w' and w_1' respectively, the

extra displacement at either end of the beam due to the yielding of the foundation columns is

$$\left. \begin{aligned} w' + 3w_1'L &= W[18(\mu L)^2 + 6(\mu L) + 1]/2\mu^3 EI \\ &= WL^3 q[18 + 6q + q^2]/2EI. \end{aligned} \right\} \quad (7.3)$$

This is exactly the same as the expression for δ in equation (4.5), which is the extra displacement at the end of a cantilever which projects $3L$ beyond a foundation of length L . The expression (7.3) makes it clear that the same extra displacement arises for a beam of the same length in foundations of half the length, even though the boundary conditions at $x=0$ differ slightly. In this section the symmetry requires $\frac{dy}{dx}=0$ at

$x=0$, whereas in section 4 the corresponding condition gives $\frac{d^2y}{dx^2}=0$. In each case $\frac{d^3y}{dx^3}=0$, the shearing force being zero. That (7.3) and (4.5)

coincide in form, then, is due to the fact that in each case the negative exponentials are neglected in comparison with their reciprocals, and that approximate formulæ are used. Fig. 3 shows that y and its first three derivatives are all very small at $x=0$, and thus the use of the approximate formulæ masks the slight difference in displacement, obtained by equating what is a very small slope to zero.

Suppose now, that with $\beta=1$, the loading at one of the free ends, say $x=-L/2$, is reversed in direction. The same solution is obtained as in the previous case for the displacement within the foundation. The disposition of the loading now requires that $y(x)=-y(-x)$, which implies that $C=-A$, and $B=D$.

Neglecting the negative exponential terms, as before, the approximate form for y may be taken, viz.,

$$y = \exp(\mu x)[A \cos \mu x + B \sin \mu x].$$

The usual boundary conditions at $x=L/2$ yield the following equations for A and B :

$$3WL/2\mu^2 EI = \exp(\mu L/2) \left[B \cos \frac{\mu L}{2} - A \sin \frac{\mu L}{2} \right],$$

$$\text{and} \quad W/2\mu^3 EI = \exp(\mu L/2) \left[(A-B) \cos \frac{\mu L}{2} + (A+B) \sin \frac{\mu L}{2} \right].$$

The values of w_1' and w' are found in a similar manner to that employed above, to be given by

$$w' = W(3\mu L + 1)/2\mu^3 EI,$$

$$\text{and} \quad w_1' = W(6\mu L + 1)/2\mu^2 EI.$$

Thus δ , the extra deflexion at the free end due to the elastic nature of the foundation columns, is given by

$$\delta = w' + 3w_1'L = WL^3 q(18 + 6q + q^2)/2EI,$$

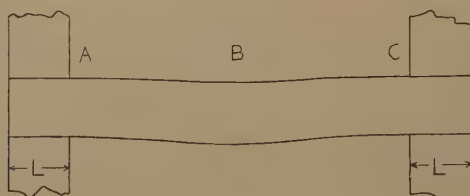
the same result as (7.3). This might have been anticipated, since reversing the direction of the load at one of the free ends will merely reverse the displacements in that half of the beam.

8. *Beam held at both Ends within Elastic Foundations.*

Finally, a beam is supposed to be held at each end between foundation columns of length L , as shown in fig. 7, and to be subject to some system of loading between A and C which is symmetrical about the mid-point B.

The additional displacements at its mid-point B, due to the yielding of the foundation columns, may be calculated by finding first the bending moment¹ and shearing force at A or C. On the assumption that, with $\beta=1$, the dimensions involved satisfy the inequality (4.4), then the approximate forms for slope and displacement may be used, and the extra displacement δ , at B, may be calculated by using equation (4.6)

Fig. 7.



Beam supported at both ends.

If the loading of the beam is not symmetrical about B, then the shearing force and bending moments at C and A will differ. The extra deflexion δ at B will then be given by

$$\delta = (\delta_1 + \delta_2)/2,$$

where δ_1 is calculated from the values of W and M at A, and δ_2 from those at C.

XXI. *Magnetic Moments and the Vibratory Electron.*

By IAN A. ROBERTSON, M.A., B.Sc.,
Research Student in the University of Glasgow *.

[Received July 31, 1942.]

THE problem of the magnetic moments of revolving electron systems has recently been investigated by means of the vibratory electron (Phil. Mag. vol. xxv. pp. 682-693 (1938)). It was then shown that the usual formula for such systems, given by the current multiplied by the area in which the electrons revolve, does not hold for two pairs of electrons, revolving at diametrically opposite points in coaxial circles about two fixed central positive electrons. On the other hand, the vibrational theory agrees with Ampère's theorem concerning the equivalence of an

* Communicated by Professor E. Taylor Jones.

electric circuit and a magnetic shell, having the circuit as the bounding edge, provided the number of the electrons in the system is large. In the present communication, this discrepancy between the magnetic moments of systems of revolving electrons, and Ampère's theorem for electrical currents is investigated. It would appear that the magnetic moment is not an intrinsic property of revolving electron systems but depends to a large extent upon the orientation of neighbouring electron systems. In the case of a negative electron, revolving in a circle around a fixed positive electron—a system similar to the Bohr magneton and the hydrogen atom—it will be shown that this system has properties differing from all the other systems of revolving electrons. Indeed the system of one revolving negative electron and one fixed positive electron, when coupled with a similar system has really no simple magnetic moment, but there is an average effect somewhat similar to that of a magnetic moment.

When two or more pairs of electrons are revolving in coaxial circles about fixed central positive charges, simple magnetic moments exist; but for two and three pairs of electrons the values of these magnetic moments differ from the usual values. However, for four or more electrons, when coupled with a similar system, the vibrational theory agrees with the formula derived from Ampère's theorem.

These results may be of interest in the explanation of various atomic phenomena.

Summary of some of the Properties of the Vibratory Electron.*

The electron is to be regarded both when in motion and when at rest as a simple vibrator, having the frequency of the oscillations surrounding it. When isolated, the electron has neither electrical nor magnetic energy, but acquires one or both when its frequency is changed by the "coupling" effects of other electrons upon it.

One of the fundamental equations of the vibratory electron theory is derived by considering two electrons, or two groups of similarly situated electrons, when they become "coupled" with each other. If the square of the periodicity of the first group when separate is $a_1 = 4\pi\nu_1^2$, its equation of motion will be $\ddot{x} + a_1x = 0$, and similarly the equation for the second group is

$$\ddot{z} + a_2z = 0.$$

If the coupling of the first group with that of the second, when they are no longer isolated is b_1 and *vice versa* b_2 , then we have

$$\ddot{x} + a_1x + b_1z = 0,$$

$$\ddot{z} + a_2z + b_2x = 0.$$

Assuming $x = Ae^{int}$, $z = Be^{int}$, where $n/2\pi$ is the common frequency of the two groups, we have

$$n^2 = \frac{a_1 + a_2}{2} \pm \frac{1}{2} \sqrt{(a_1 - a_2)^2 + 4b_1b_2}. \quad \dots \quad (1)$$

* Phil. Mag. vol. xxi. pp. 337-355 (1936); vol. xxii. pp. 921-938 (1936).

and
$$\frac{A}{B} = \frac{n^2 - a_2}{b_2} \dots \dots \dots (1a)$$

A special case of the above occurs when

$$a_1 = a_2 = a, \text{ and } b_1 = b_2 = b.$$

Then
$$n^2 = a + b \dots \dots \dots (2)$$

or
$$n^2 = a - b \dots \dots \dots (3)$$

From these last two equations, it can be shown that electrons at rest and at moderate distances apart obey the usual electrostatic laws and each electron has a constant charge e given by $e^2 = \frac{hk}{2\pi\sqrt{a}}$, where h = Planck's constant, k = constant, $a = 4\pi^2\nu^2$, ν being the frequency of the electron. At rest a free electron has the frequency $\nu = 1.24 \times 10^{20}$.

Also $b = \frac{hk}{r}$, where r is distance between the electrons.

However, the more interesting case for the vibratory electron theory occurs when two electrons approach very close to each other.

As the electrons approach, the "coupling" b increases, but its greatest possible value is a . This value is attained when the electrons are about 2.8×10^{-13} cm. apart.

When b has attained this value, it can be seen from equations (2) and (3) that

$$n^2 = 2a \text{ or } 0.$$

In the first case there are two similar electrons, so close together that there is no longer any repulsion between them. In the second case the electrons annihilate each other or appear as a photon of light*.

Considerable use will be made in what follows of electrons within this critical distance of each other, where no repulsion takes place.

Magnetic Terms.

Magnetic forces and energy can also be explained by the vibratory electron †.

It is assumed that the magnetic terms depend upon the translational velocities of the electrons, and that an electron at rest has no magnetic energy. The velocities of all the electrons in this communication will be assumed small compared to that of light, hence, all relativity corrections can be neglected. The electrons will be considered in moving pairs and we have two cases to consider.

1. The velocities v_1 and v_2 of the two moving electrons, and the line r joining the electrons, all lie in the same plane (fig. 1).

The angles made by v_1 and v_2 with r are θ_1 and θ_2 respectively.

* Phil. Mag. vol. xxiv. pp. 458-466 (1937).

† Phil. Mag. vol. xxv. pp. 682-693 (1938).

Values of the total couplings b_1 and b_2 between the two electrons have been suggested by Professor E. Taylor Jones, viz. :—

$$b_1 = b - b \frac{v_1 v_2}{c^2} \sin \theta_2,$$

$$b_2 = b - b \frac{v_1 v_2}{c^2} \sin \theta_1.$$

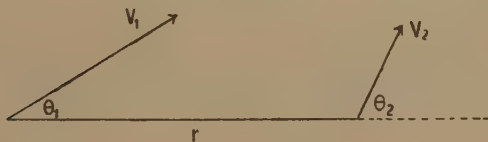
This will be applied to the fundamental frequency formula (1), and it will be assumed that the two electrons when isolated have equal periodicities, i. e., $a_1 = a_2$.

Therefore

$$\begin{aligned} n^2 &= a \pm \frac{1}{2} \left[4 \left(b - b \frac{v_1 v_2}{c^2} \sin \theta_2 \right) \left(b - b \frac{v_1 v_2}{c^2} \sin \theta_1 \right) \right]^{\frac{1}{2}} \\ &= a \pm b \left[1 - \frac{v_1 v_2}{c^2} (\sin \theta_1 + \sin \theta_2) + \left(\frac{v_1 v_2}{c^2} \right)^2 \sin \theta_1 \sin \theta_2 \right]^{\frac{1}{2}}. \end{aligned}$$

As v_1 and v_2 are small compared to c , $\left(\frac{v_1 v_2}{c^2} \right)^2$ can be neglected, and the

Fig. 1.



expression under the root sign can be expanded by the binomial theorem, giving

$$n^2 = a \pm b \left[1 - \frac{1}{2} \frac{v_1 v_2}{c^2} (\sin \theta_1 + \sin \theta_2) \right]. \quad (4)$$

From equation (2) we have that the common frequency $n/2\pi$ for two similar stationary electrons is given by $n^2 = a + b$. Hence the magnetic coupling of two similar moving electrons will be

$$Lv_1 v_2 = -\frac{1}{2} b \frac{v_1 v_2}{c^2} (\sin \theta_1 + \sin \theta_2). \quad (5)$$

This formula is fundamental in all magnetic terms.

2. The velocities v_1 and v_2 of the two moving electrons and the line r joining them, do not lie in the same plane.

In this case v_1 is resolved orthogonally into the plane containing v_2 and r , and the resolved part is treated with v_2 as in the previous case. Thus if v_1 is perpendicular to the plane of v_2 and r there will be no magnetic but only electrostatic coupling. This last statement would appear to be in keeping with the observed fact that there is no magnetic attraction between two mutually perpendicular wires carrying an electric current.

Magnetic Moment of Simple Circuits.

The first system here considered will consist of two coaxial circles distance s apart : each circle has a fixed central positive electron and a negative electron revolving around its circumference with velocity v . The radius of each circle is r , and r is small compared to s , so that powers higher than $\frac{r^2}{s^2}$ can be neglected (fig. 2).

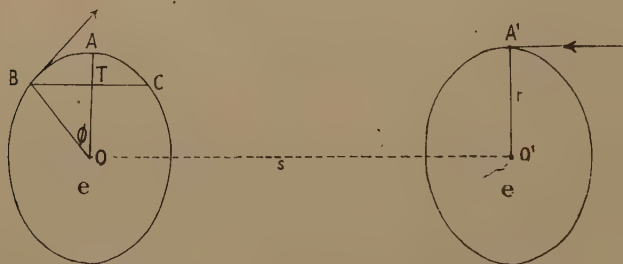
At a certain instant, the negative electrons are at B and A' : $\hat{BOA} = \phi$ and AO is parallel to O'A'.

The system will be divided into two groups, viz. :—

Group I. The two central electrons at O and O'.

Group II. The two revolving electrons, situated instantaneously at B and A'.

Fig. 2.



Two parallel circular orbits with phase difference ϕ , with electrons at B and A'.

All the electrons of the first group are similarly situated, in regard to sign as well as position, with respect to those of the second group.

Hence

$$a_1 = a + k/s$$

$$a_2 = a + k/y + Lv^2,$$

where $y = A'B$ and Lv^2 is magnetic coupling of the electrons at A' and B.

The square of the periodicity of each electron (both positive and negative) when isolated is a , of the first group when isolated a_1 , and of the second a_2 . The coupling of each electron of the first group with those of the second is b_1 , and similarly the coupling of each electron of the second group with all those of the first group is b_2 .

where

$$b_1 = \frac{k}{r} + \frac{k}{x} = b_2 \text{ and } x = O'B.$$

In the frequency formula (1) for two groups of electrons the negative sign is taken when the two groups of electrons are of different sign. Therefore

$$n^2 = a + \frac{1}{2} \left(\frac{k}{s} + \frac{k}{y} \right) + \frac{1}{2} Lv^2 - \frac{1}{2} \left[\left(\frac{k}{s} - \frac{k}{y} - Lv^2 \right)^2 + 4b_1^2 \right]^{\frac{1}{2}}.$$

The maximum value of $\left(\frac{k}{s} - \frac{k}{y}\right)$ is $\frac{2kr^2}{s^3}$, and the maximum value of the magnetic coupling $Lv^2 = -\frac{kv^2}{sc^2}$.

To the order of magnitude already stated,

$$\begin{aligned} n^2 &= a + \frac{1}{2} \left(\frac{k}{s} + \frac{k}{y} \right) + \frac{1}{2} Lv^2 - b_1 \\ &= a + \frac{k}{2} \left(\frac{1}{s} + \frac{1}{y} - \frac{1}{r} - \frac{1}{x} \right) + \frac{1}{2} Lv^2. \end{aligned}$$

Now a is much greater than all the other terms, therefore

$$\begin{aligned} n &= \sqrt{a} + \frac{1}{2\sqrt{a}} \left[\frac{k}{2} \left(\frac{1}{s} + \frac{1}{y} - \frac{1}{r} - \frac{1}{x} \right) + \frac{1}{2} Lv^2 \right] \\ &= 2\pi\nu. \end{aligned}$$

If s is infinite, we have

$$\begin{aligned} n_1 &= \sqrt{a} - \frac{1}{2\sqrt{a}} \left(\frac{k}{2} \cdot \frac{1}{r} \right) \\ &= 2\pi\nu_0. \end{aligned}$$

Therefore $n - n_1 = 2\pi(\nu - \nu_0)$

$$= \frac{1}{2\sqrt{a}} \left[\frac{k}{2} \left(\frac{1}{s} + \frac{1}{y} - \frac{1}{x} \right) + \frac{1}{2} Lv^2 \right].$$

The mutual energy of the system, when all the four electrons are oscillating with the frequency ν , is from elementary quantum theory $4h(\nu - \nu_0)$. Therefore

$$4h(\nu - \nu_0) = \frac{h}{\pi\sqrt{a}} \left[\frac{k}{2} \left(\frac{1}{s} + \frac{1}{y} - \frac{1}{x} \right) + \frac{1}{2} Lv^2 \right].$$

The ratio of the amplitudes $\frac{A}{B}$ of the vibration of the first and second groups is, by (1 a),

$$\frac{A}{B} = \frac{n^2 - a_2}{b_2}, \text{ which is not zero.}$$

The central particles, therefore, in this case, do take part in the oscillation, and thus contribute to the magnetic energy. The magnetic energy is given by the expression

$$E_m = \frac{1}{2} \frac{h}{\pi\sqrt{a}} Lv^2.$$

The expression for the magnetic coupling Lv^2 has now to be found.

The velocities at A' and B and the line A'B do not lie in the same plane. The line BC (fig. 2) which is perpendicular to OA and hence parallel to the velocity at A', lies in the plane of the velocity at A' and the line A'B. The velocity at B can now be resolved parallel and perpendicular to BC, giving $v \cos \phi$ and $v \sin \phi$. Since $v \cos \phi$ and the

velocity of A' now lie in the same plane through the line joining the particles, the magnetic coupling formula (5) can be applied, viz. :—

$$Lv_1v_2 = -\frac{1}{2}b \frac{v_1v_2}{c^2} (\sin \theta_1 + \sin \theta_2), \dots$$

with $v_1 = v \cos \phi$, $v_2 = v$; $\theta_1 = \theta_2$,

and $b = \frac{k}{A'B} = \frac{k}{y}$.

Therefore $L_1v^2 = -b \frac{v^2}{c^2} \cos \phi \sin \theta_1$, where θ_1 is the angle between A'B and BT (fig. 2) and

$$\hat{BOT} = \phi.$$

Now $BT = r \sin \phi$; $OB = r$; $TO = r \cos \phi$;

$$A'T = [s^2 + r^2(1 - \cos \phi)^2]^{\frac{1}{2}};$$

and $y = A'B = [s^2 + r^2(1 - \cos \phi)^2 + r^2 \sin^2 \phi]^{\frac{1}{2}}.$

Hence $\sin \theta_1 = \frac{[s^2 + r^2(1 - \cos \phi)^2]^{\frac{1}{2}}}{[s^2 + r^2(1 - \cos \phi)^2 + r^2 \sin^2 \phi]^{\frac{1}{2}}}.$

Therefore $L_1v^2 = -k \frac{v^2}{c^2} \cos \phi \frac{[s^2 + r^2(1 - \cos \phi)^2]^{\frac{1}{2}}}{[s^2 + r^2(1 - \cos \phi)^2 + r^2 \sin^2 \phi]^{\frac{1}{2}}}$

$$= -\frac{kv^2}{sc^2} \cos \phi \frac{\left[1 + \frac{r^2}{s^2}(1 - \cos \phi)^2\right]^{\frac{1}{2}}}{\left[1 + \frac{r^2}{s^2}(1 - \cos \phi)^2 + \frac{r^2}{s^2} \sin^2 \phi\right]}.$$

Expanding and neglecting powers higher than $\frac{r^2}{s^2}$, we have

$$L_1v^2 = -\frac{kv^2}{c^2s} \cos \phi + \frac{kv^2r^2}{s^3c^2} \left[\frac{\cos \phi}{2} [1 - \cos \phi]^2 + \cos \phi \sin^2 \phi \right]. \quad (6)$$

The magnetic coupling L_2v^2 for the velocities $v \sin \phi$ and v at B and A' respectively, will now be considered. Referring again to fig. 2, we see that $v \sin \phi$ and v are not in the plane containing y , but $v \sin \phi$ can be resolved parallel and perpendicular to TA' in the plane of y and velocity v at A', giving $v \sin \phi \cos \psi$ and $v \sin \phi \sin \psi$, where ψ is equal to the angle ATA' in fig. 2. Now $v \sin \phi \sin \psi$ is perpendicular to the velocity at A' and to the plane containing the velocity at A' and the line y , and it has already been assumed that such terms have no magnetic coupling. The magnetic coupling for $v \sin \phi \cos \psi$ at B and v at A' remains to be considered.

Applying the magnetic coupling formula (5)

$$L_2v^2 = -\frac{1}{2}b \frac{v \sin \phi \cos \psi \cdot v}{c^2} [\sin \theta_1' + \sin \theta_2'], \quad (7)$$

where $b = \frac{k}{y} = \frac{k}{A'B}$; and as $v \sin \phi \cos \psi$ and v at A' are perpendicular,

we have $\theta_1' = \frac{\pi}{2} - \theta_2'.$

As s , the distance between the circles, is large compared to r , the radius of each circle, $\cos \psi$ is always small, its maximum value being $\frac{2r}{s}$.

Hence L_2 is of order $\frac{r}{s^2}$.

The total magnetic coupling of the two electrons is $Lv^2 = (L_1 + L_2)v^2$, and when $\phi = 0$ this becomes

$$Lv^2 = -\frac{kv^2}{c^2s}.$$

The magnetic energy E_m is $\frac{1}{2} \frac{hk}{\pi\sqrt{a}} Lv^2$, i. e. when $\phi = 0$, $E_m = -\frac{1}{2} \frac{hk}{\pi\sqrt{a}} \frac{v^2}{c^2s}$.

Now $e^2 = \frac{1}{2} \frac{hk}{\pi\sqrt{a}} \cdot \frac{1}{c^2}$ in e.m. units, so that when $\phi = 0$, $E_m = -\frac{e^2v^2}{s}$, and the mutual attraction F_m , or $\frac{dE_m}{ds}$, is e^2v^2/s^2 . But the mutual attraction of two small electrical circuits, each of magnetic moment M , placed coaxially at distance s apart, is $6M^2/s^4$.

Evidently these two attractions F_m and $6M^2/s^4$ cannot be equal, for if they were equated, M^2 would be directly proportional to s^2 , a result which is obviously wrong. The vibratory electron theory would appear to indicate that for two such simple systems when coupled with the moving electrons in phase, i. e. with $\phi = 0$, there is no ordinary magnetic moment. It can also be readily shown by considering the magnetic coupling Lv^2 for all values of ϕ , that at no given position of the two moving electrons is there any definite magnetic moment. However, as ϕ varies from 0 to 2π the mean magnetic coupling effect can be found.

The mean magnetic coupling is

$$\begin{aligned} \overline{Lv^2} &= \frac{v^2}{2\pi} \int_0^{2\pi} L d\theta = \frac{v^2}{2\pi} \int_0^{2\pi} (L_1 + L_2) d\theta \\ &= -\frac{kv^2r^2}{2s^3c^2}, \text{ since } \int_0^{2\pi} L_2 d\theta = 0. \end{aligned}$$

$$\begin{aligned} \text{Thus } \overline{E_m} &= -\frac{1}{2} \frac{hk}{\pi\sqrt{a}} \cdot \frac{r^2v^2}{2s^3c^2} \\ &= -\frac{e^2r^2v^2}{2s^3}, \end{aligned}$$

and the force of attraction between the two circuits is

$$\overline{F_m} = \frac{d\overline{E_m}}{ds} = \frac{3e^2r^2v^2}{s^4}.$$

Equating this to $6M^2/s^4$ we find $M = \frac{1}{2} evr$.

Now the expression for the magnetic moment derived from Ampère's formula is $M = \frac{1}{2} Ne vr$, where N is the number of revolving electrons in each circle. In the system under consideration $N = 1$. Hence the mean effect calculated from the vibratory theory and Ampère's formula agree. No doubt, when there is a very large number of such systems, the average effect will predominate, yet it does not seem unlikely that in certain cases the moving electrons will keep their positions with respect to each

other, and it will then be wrong to assume that the magnetic moment is given by $\frac{1}{2}evr$, or even to assume that a magnetic moment exists for such systems.

The next system to be considered will consist of two coaxial circles distance s apart. Each circle has two fixed central positive electrons within the critical distance of each other and a negative electron revolving around its circumference with velocity v .

The reasoning is similar to the previous case.

$$a_1 = 2a + \frac{2k}{s}.$$

$$a_2 = a + \frac{k}{y} + Lv^2.$$

$$b_1 = \frac{k}{r} + \frac{k}{x}.$$

$$b_2 = \frac{2k}{r} + \frac{2k}{x}.$$

By formula (1) the common frequency n is given by

$$n^2 = \frac{a_1 + a_2}{2} - \left((a_1 - a_2)^2 + 4b_1b_2 \right)^{\frac{1}{2}},$$

and
$$(a_1 - a_2)^2 = \left(a + \frac{2k}{s} - \frac{k}{y} - Lv^2 \right)^2;$$

as a is much larger than the other terms $(a_1 - a_2)^2$ is much larger than $4b_1b_2$. Hence

$$n^2 = a_2 - \frac{b_1b_2}{a_1 - a_2},$$

so that, to the order of magnitude stated, we have

$$n^2 - a_2 = 0.$$

Now, by formula 1(a),
$$\frac{A}{B} = \frac{n^2 - a_2}{b_2}.$$

Hence $A=0$. Thus the central particles do not take part in the oscillation of frequency ν , and do not contribute to the magnetic energy of the revolving electrons. It might be pointed out here, that no matter how many positive electrons there are at the centre of each circle, $(a_1 - a_2)^2$ will always be very much larger than $4b_1b_2$ and $n^2 - a_2 = 0$, except in the case already discussed where there is only one positive electron at the centre.

Now
$$n^2 = a_2 = a + \frac{k}{y} + Lv^2 = 4\pi^2\nu^2.$$

Hence
$$n = \sqrt{a} + \frac{1}{2} \frac{1}{\sqrt{a}} \frac{k}{y} + \frac{1}{2} \frac{1}{\sqrt{a}} Lv^2 = 2\pi\nu,$$

and as
$$s \rightarrow \infty \quad n \rightarrow n_0 \rightarrow \sqrt{a} = 2\pi\nu_0,$$

$$\therefore 2\pi(\nu - \nu_0) = \frac{1}{2} \frac{1}{\sqrt{a}} \frac{k}{y} + \frac{1}{2} \frac{1}{\sqrt{a}} Lv^2.$$

The total energy of the system is from elementary quantum theory $2h(\nu-\nu_0)$ and

$$2h(\nu-\nu_0) = \frac{1}{2} \frac{1}{\pi\sqrt{a}} \frac{k}{y} + \frac{1}{2} \frac{1}{\pi\sqrt{a}} Lv^2.$$

Hence the magnetic energy is $E_m = \frac{1}{2\pi\sqrt{a}} Lv^2$.

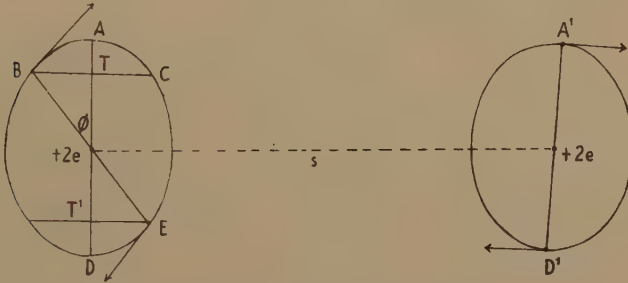
This is the same value for the magnetic energy as for the previous case, when there is only one electron at the centre. The average magnetic coupling is also the same, viz., $\bar{Lv}^2 = -\frac{kv^2r^2}{2s^3}$.

Therefore, $\bar{E}_m = -\frac{1}{2} \frac{e^2v^2r^2}{s^3},$

and $\bar{F}_m = \frac{d\bar{E}_m}{ds} = \frac{3}{2} \frac{e^2v^2r^2}{s^3} = \frac{6M^2}{s^4}.$

Therefore, $M = \frac{1}{2} evr.$

Fig. 3.



Two parallel circular orbits with phase difference ϕ .

The next problem to be considered was first discussed on page 686, *Phil. Mag.* xxv. (1938).

It was a system of two pairs of electrons, revolving at diametrically opposite points, in coaxial circles, about two fixed central positive charges, but the diameters on which the electrons are placed, instead of being parallel, will now be inclined to each other at the angle ϕ (fig. 3).

By a method similar to that discussed in the previous problem, we find that the mutual energy of the two revolving systems, if all the electrons are oscillating with the frequency ν , is

$$8h(\nu-\nu_0) = \frac{2hk}{\pi\sqrt{a}} \left(\frac{1}{y} + \frac{1}{z} \right) + \frac{2hLv^2}{\pi\sqrt{a}},$$

where $y = A'B$, and $z = A'D$,

and Lv^2 is the magnetic coupling of the electron at A' with those at B and E .

From (6) the magnetic coupling of the electrons at A' and B is

$$L_1v^2 = -\frac{kv^2}{c^2s} \cos \phi + \frac{kv^2r^2}{s^3c^2} \left[\frac{\cos \phi}{2} [1 - \cos \phi]^2 + \cos \phi \sin^2 \phi \right] + L_2v^2,$$

where $L_2 v^2$ is coupling for velocity v at A' with the "perpendicular" component velocity $v \sin \phi$ at B .

Similarly, the magnetic coupling of A' and E is, when $\phi' = \phi + \pi$,

$$L_1' v^2 = \frac{kv^2}{c^2 s} \cos \phi + \frac{kv^2 r^2}{s^3 c^2} \left[-\frac{\cos \phi}{2} (1 + \cos \phi)^2 - \cos \phi \sin^2 \phi \right] + L_2' v^2.$$

The total coupling $L v^2 = (L_1 + L_1') v^2$

$$= \frac{kv^2 r^2}{s^3 c^2} [-\cos^2 \phi] + (L_2 + L_2') v^2.$$

We now have to consider the magnetic couplings $L_2 v^2 + L_2' v^2$. $L_2 v^2$ is the magnetic coupling of $v \sin \phi$ at B with v at A' , and $L_2' v^2$ is the magnetic coupling of $-v \sin \phi$ at E with v at A' .

The magnetic coupling of L_2 has already been discussed in the previous case (viz. equation (7)),

$$L_2 v^2 = -\frac{1}{2} b \frac{v \sin \phi \cos \psi \cdot v}{c^2} [\sin \theta_1' + \sin \theta_2'], \quad \dots \quad (7)$$

and L_2' is treated similarly,

$$L_2' v^2 = +\frac{1}{2} b' \frac{v \sin \phi \cos \psi' \cdot v}{c^2} [\sin \theta_1'' + \sin \theta_2''], \quad \dots \quad (8)$$

where ψ' is angle $AT'A$, and $\theta_1'' = \frac{\pi}{2} - \theta_2''$.

$$b' = \frac{k}{A'E},$$

as s is large compared to r , b is nearly equal to b' and maximum value of $|\sin \theta_1' + \sin \theta_2'| = \sqrt{2}$.

Therefore, $(L_2 + L_2') v^2 \leq \frac{1}{2} b \frac{v^2 \sin \phi}{c^2} \sqrt{2} (\cos \psi' - \cos \psi)$,

and as ψ' is nearly equal to ψ , $L_2 + L_2'$ can be neglected.

Hence
$$L v^2 = -\frac{2kv^2 r^2}{s^3 c^2} \cos^2 \phi.$$

It has already been shown that the central particles do not contribute to the energy of the revolving electrons, therefore the mutual energy of the system depends upon four and not eight revolving electrons, and is equal to

$$4h(v - v_0) = \frac{hk}{\pi\sqrt{a}} \left(\frac{1}{y} + \frac{1}{z} \right) + \frac{h}{\pi\sqrt{a}} L v^2.$$

The magnetic energy is
$$E_m = -\frac{2h \cdot kv^2 r^2}{\pi\sqrt{a} s^3 c^2} \cos^2 \phi,$$

but

$$e^2 = \frac{hk}{2\pi\sqrt{a}} \cdot \frac{1}{c^2} \text{ in electromagnetic units,}$$

$$\therefore E_m = -\frac{4e^2 v^2 r^2 \cos^2 \phi}{s^3}.$$

The mutual magnetic attraction is F_m and

$$F_m = \frac{dE_m}{ds} = \frac{12e^2 v^2 r^2 \cos^2 \phi}{s^4}.$$

The mutual attraction of two small circuits of magnetic moment M distance s apart is

$$\begin{aligned} F_m &= \frac{6M^2}{s^4}, \\ \therefore \frac{6M^2}{s^4} &= \frac{12e^2v^2r^2}{s^4} \cos^2 \phi, \\ M^2 &= 2e^2v^2r^2 \cos^2 \phi, \\ M &= \sqrt{2} |\cos \phi| evr. \end{aligned}$$

The mean value of the magnetic moment as ϕ varies from 0 to π is

$$\bar{M} = \frac{1}{\pi} \int_0^\pi |\cos \phi| d\phi \sqrt{2} evr = 0.9003 evr.$$

The usual formula for the magnetic moment is $\frac{1}{2}Nerv$, where N is number of revolving electrons.

In the present case $N=2$, so the old formula for the magnetic moment of such a system is

$$M = erv,$$

while the vibrational mean value formula is

$$\bar{M} = 0.9003 evr,$$

a difference of almost 10 per cent. between the two values.

It will be noted that when $\phi=0$

$$M = \sqrt{2} evr$$

and when

$$\phi = \frac{\pi}{2}$$

$$M = 0.$$

Thus it would appear that the magnetic moment of a system of revolving electrons is not an intrinsic property, but depends upon the external circuits.

The next case to be considered is that of three pairs of electrons, moving around two coaxial circles, distance S apart, having three fixed positive central electrons within critical distance of each other (fig. 4). The electrons on each circle are at the vertices of an equilateral triangle.

$\triangle BEF$ and $\triangle A'B'C'$ are two equilateral triangles inscribed in two coaxial circles, having centres O and O' respectively. Each circle is of radius r .

OB is inclined to OA at the angle ϕ and OA is parallel to $O'A'$.

With reasoning similar to the previous cases, we find that the central particles do not take part in the oscillation but vibrate as an independent system, and that the magnetic energy is given by

$$E_m = \frac{3}{2\pi} \frac{h}{\sqrt{a}} Lv^2.$$

The problem is now reduced to finding the magnetic coupling between one of the revolving electrons of one circle and the three electrons of the other circle. The method is simply a further application of the formula already found for the magnetic coupling, viz. (6),

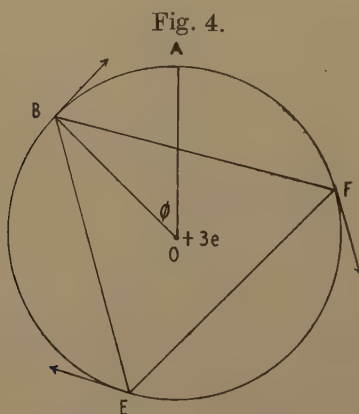
$$L_1 v^2 = -\frac{kv^2}{c^2 s} \cos \phi + \frac{kv^2}{s^3 c^2} r^2 \left[\frac{\cos \phi}{2} [1 - \cos \phi]^2 + \cos \phi \sin^2 \phi \right], \quad (6)$$

but in this problem we have to consider the coupling for the angles—

$$\phi, \quad \phi + \frac{2\pi}{3}; \quad \phi + \frac{4\pi}{3};$$

now

$$\begin{aligned} \cos \phi + \cos \left(\phi + \frac{2\pi}{3} \right) + \cos \left(\phi + \frac{4\pi}{3} \right) &= 0, \\ \cos^2 \phi + \cos^2 \left(\phi + \frac{2\pi}{3} \right) + \cos^2 \left(\phi + \frac{4\pi}{3} \right) &= \frac{3}{2}, \\ \cos^3 \phi + \cos^3 \left(\phi + \frac{2\pi}{3} \right) + \cos^3 \left(\phi + \frac{4\pi}{3} \right) &= \frac{3}{4} \cos 3\phi \end{aligned}$$



Cross-section diagram of two parallel circular orbits, with phase difference ϕ . Each orbit has three electrons symmetrically placed.

The total magnetic coupling $Lv^2 = (L_1 + L_2 + L_3)v^2$, so the total coupling

$$Lv^2 = -\frac{3}{2} \frac{kv^2 r^2}{s^3 c^2} \left[1 + \frac{1}{4} \cos 3\phi \right].$$

Therefore

$$\begin{aligned} E_m &= -\frac{9}{4} \frac{hkv^2 r^2}{s^3 c^2} \left[1 + \frac{1}{4} \cos 3\phi \right] \\ &= -\frac{9}{2} \frac{e^2 v^2 r^2}{s^3} \left[1 + \frac{1}{4} \cos 3\phi \right]. \end{aligned}$$

The force of attraction

$$F_m = \frac{dE_m}{ds} = +\frac{9 \cdot 3}{2} \cdot \frac{e^2 v^2 r^2}{s^4} \left[1 + \frac{1}{4} \cos 3\phi \right] = \frac{6M^2}{s^4}.$$

Hence

$$M = \frac{3}{2} e v r \left[1 + \frac{1}{4} \cos 3\phi \right]^{\frac{1}{2}}.$$

The mean value for M as θ varies from θ to $\frac{2\pi}{3}$ is

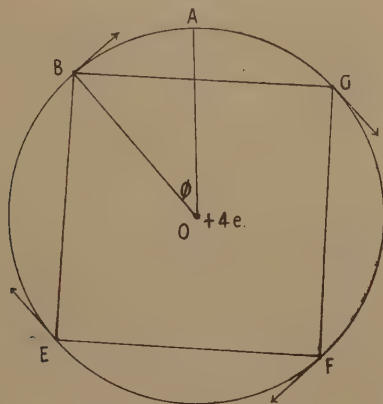
$$\begin{aligned}\bar{M} &= \frac{3}{2} evr \left(\frac{3}{2\pi} \right) \int_0^{\frac{2\pi}{3}} \left(1 + \frac{1}{4} \cos 3\phi \right) d\phi \\ &= \frac{3}{2} evr \cdot 0.997. \\ \bar{M} &= 1.4955 evr.\end{aligned}$$

This is again different from the usual value given by $\frac{1}{2} N evr$, where $N=3$ and the magnetic moment given by Ampère's formula is $1.5 evr$.

In this problem, as in the last, the magnetic coupling of the "perpendicular components" terms cancels out.

The problem of four electrons on each circle will now be considered.

Fig. 5.



Cross-section diagram of two parallel circular orbits with phase difference ϕ . Each orbit has four electrons symmetrically arranged.

The conditions are similar to those of the previous problem, and the magnetic energy in this case is

$$E_m = \frac{2\hbar}{\pi\sqrt{a}} L v^2.$$

The electrons are arranged in the circles at the vertices of squares BGFE and A'B'C'D' (fig. 5).

OB is inclined at angle ϕ to OA, and OA is parallel to O'A'. The magnetic coupling $L v^2$ is given by (6), the angles required being

$$\phi, \quad \phi + \pi/2, \quad \phi + \pi, \quad \phi + 3\pi/2.$$

The total coupling $L v^2$, i. e. $(L_1 + L_2 + L_3 + L_4) v^2$ is

$$\begin{aligned}L v^2 &= -\frac{kv^2}{s} \sum_{R=0}^3 \cos \left(\phi + R \frac{\pi}{2} \right) + \frac{kv^2 r^2}{s^3} \sum_{R=0}^3 \left[\frac{\cos \phi'}{2} (1 - \cos \phi')^2 + \cos \phi' \sin^2 \phi' \right]^* \\ &= -2kv^2 r^2 / s^3.\end{aligned}$$

* In this expression ϕ' is written for $\phi + R \frac{\pi}{2}$.

Thus

$$E_m = -4hkv^2r^2/\pi s^3\sqrt{a} \\ = -8e^2v^2r^2/s^3,$$

and

$$F_m = dE_m/ds = 24e^2v^2r^2/s^4 = 6M^2/s^4,$$

so that

$$M = 2evr.$$

In this case there is agreement between Ampère's formula and the vibrational theory.

When there are more than four electrons arranged regularly in such systems of coaxial circles as we have considered, we find that the vibrational theory gives the same results for the magnetic moment as that given by Ampère's formula.

In general the magnetic energy is given by

$$E_m = \frac{n}{2} \frac{h}{\pi\sqrt{a}} Lv^2,$$

where n is number of electrons revolving in each circle around the same number of fixed positive central electrons,

$$L = \sum_{i=0}^n L_i, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

now in formula (6).

$$\sum_{R=0}^{n-1} \cos\left(\phi + R\frac{2\pi}{n}\right) \equiv 0, \\ \sum_{R=0}^{n-1} \cos^2\left(\phi + R\frac{2\pi}{n}\right) = \frac{n}{2} \text{ for } n > 3, \\ \sum_{R=0}^{n-1} \cos^3\left(\phi + R\frac{2\pi}{n}\right) \equiv 0 \text{ for } n > 3.$$

Hence

$$L = \sum_{i=0}^n L_i = -\frac{kv^2r^2}{s^3} \cdot \frac{n}{2}.$$

Therefore,

$$E_m = -\left(\frac{n}{2}\right)^2 \frac{2e^2v^2r^2}{s^3}, \\ F_m = +\frac{dE_m}{ds}, \\ F_m = \left(\frac{n}{2}\right)^2 \frac{6e^2v^2r^2}{s^4} = \frac{6M^2}{s^4}. \\ \therefore M = \frac{n}{2} evr,$$

where n is number of revolving electrons.

This result is in complete agreement with Ampère's formula.

I wish to express my thanks to Professor E. Taylor Jones for suggesting these problems and for his encouragement and advice. I also wish to thank the Glasgow University for the facilities granted to me to undertake this research.

XXII. *Rational Electrodynamics.*—III. *The Charge as Point Singularity.*

By E. A. MILNE, F.R.S.*

[Received October 20, 1942.]

45. It is customary at the present day to begin a study of electromagnetism with the adoption of *field equations*. This would be completely at variance with the lines of thought pursued in these papers. We have seen how our formulation of gravitation arose not by the assumption of field equations, which would have been an empirical assumption, but by the analysis of the possible kinematic motions in the presence of given distributions of moving particles. We first formulated a general definition of external force, and then sought to represent this by a potential such that its gradient determined the motion. There emerged a potential χ , given by

$$\chi = -\frac{m_1 m_2 c^2}{M_0} \frac{X_{12}}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}},$$

which we identified with the gravitational potential energy of a pair of point masses, m_1 at (\mathbf{P}_1, t_1) and m_2 at (\mathbf{P}_2, t_2) , all reckoned by the specified observer O. We saw how this was equivalent to the statement of the inverse square law of gravitation in Lorentz-invariant form. We then found *a posteriori* that this solution of the gravitational problem satisfied, in t -measure, a wave equation, which can be shown to reduce, in τ -measure, simply to Laplace's equation. Field equations are thus an *a posteriori* property of the field, in our presentation, not an *a priori* assumption. Accordingly, in attempting to formulate laws of electromagnetism we must not begin with assumed field equations.

We suppose ourselves to have no empirical knowledge of electromagnetism. We put to ourselves the question: what is the next most general type of force, after gravitation, that can be abstractly formulated in the presence of the set of frames of reference we call a substratum? We seek a new kind of force associated with a singularity.

46. In both classical gravitation theory and the classical theory of electromagnetism the elementary potential possessing a singularity at $r=0$ is $1/r$. We have seen in §§ 32, 33 of the preceding paper that in our rational theory, where all observers are explicitly mentioned, the rôle of $1/r$ is played by the scalar

$$\frac{X_{12}}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}}.$$

* Communicated by the Author.

This reduces to

$$\frac{c(t_1 t_2 - \mathbf{P}_1 \cdot \mathbf{P}_2 / c^2)}{[(t_1 \mathbf{P}_2 - t_2 \mathbf{P}_1)^2 - c^{-2}(\mathbf{P}_1 \wedge \mathbf{P}_2)^2]^{\frac{1}{2}}},$$

which, being a 4-scalar, takes the same value for all fundamental observers; and this value is most readily calculated by choosing the observer at \mathbf{P}_2 , so that $\mathbf{P}_2 = 0$, when it becomes

$$\frac{ct_1}{|\mathbf{P}_1|},$$

$|\mathbf{P}_1|$ being here \mathbf{P}_2 's measure of the distance $\mathbf{P}_2 \mathbf{P}_1$ at the epoch t_1 . Since in addition $(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}$ vanishes for $\mathbf{P}_1 = \mathbf{P}_2$, $t_1 = t_2$, $(X_{12}^2 - X_1 X_2)^{-\frac{1}{2}}$ is the fundamental elementary potential possessing a singularity at the coincidence of two events.

By the addition of terms of this kind, we generate more and more complicated gravitational situations, but we shall never, by this means, emerge from the domain of gravitation. First-order partial differentiation of scalar functions of position and epoch will continue to represent only gravitational fields of force. We are seeking situations which, though still associated with a singularity at $\mathbf{P}_1 = \mathbf{P}_2$, $t_1 = t_2$, are essentially different.

47. The next hierarchy of possible forces may now be supposed derived by the *second-order* differentiation of 4-scalars which may be called super-potentials. The second-order differentiation of a scalar yields, however, a tensor of the second order, whilst what we are seeking is a vector—namely a new kind of force. To derive a vector from a tensor, our only recourse is to form its inner product with an existing vector. The only vectors we have which are associated with a particle \mathbf{P} of the substratum are the position 4-vector (\mathbf{P}, ct) and the velocity 4-vector $(\mathbf{V}/Y^{\frac{1}{2}}, c/Y^{\frac{1}{2}})$. Now for a fundamental particle we have $\mathbf{V} = \mathbf{P}/t$, and so the latter-mentioned 4-vector reduces in this case to $(\mathbf{P}/tY^{\frac{1}{2}}, ct/tY^{\frac{1}{2}})$, which, since $Y^{\frac{1}{2}}t$ is here a scalar, is just a scalar multiple of the former 4-vector. Essentially then, we have only one 4-vector at our disposal for the purpose in hand; and our new type of force is accordingly to be associated with the inner product of a tensor and the 4-vector $(\mathbf{V}/Y^{\frac{1}{2}}, c/Y^{\frac{1}{2}})$. The tensor is to be the second-order derivative of a scalar, and is to represent a force “due to” a point-singularity at \mathbf{P}_2 as measured by a test-particle \mathbf{P}_1 .

48. Any tensor of rank two can be represented as the sum of a symmetrical tensor (or self-conjugate tensor) and an anti-symmetrical tensor. Now the six distinct components of an anti-symmetrical tensor of rank 2 in four dimensions constitute a pair of 3-vectors in the associated 3-dimensional space-section of constant epoch. We shall expect these 3-vectors, if they are identifiable in nature at all, to have simple interpretations in 3-dimensional physics. I shall in this paper confine myself to the anti-symmetrical part of the general second-rank tensors which

this mode of thought suggests. The corresponding symmetrical parts can be discussed with equal ease, but they at present await physical identification in nature. The present theory is not a whit the less deductive by this procedure. *After* the deduction of the abstract scheme of relations, as in any geometry, it is necessary to consult experience to identify the abstract entities in nature; *after* the deduction, we can appeal to nature or empirical observation as much as we please; but we must be careful not to borrow from experience to eke out the deduction. That the theory about to be described throws up entities in electro-magnetism not yet identified with experimental entities is no more surprising than that pure mathematics at an elementary stage throws up complex numbers which for many years before Argand awaited physical interpretation.

49. To be explicit, then, we now seek a 4-scalar ϕ_{21} , a function of the co-ordinates, epochs, and possibly velocities of two particles P_1, P_2 , as measured by an observer O of the substratum; ϕ_{21} is to be a super-potential whose double partial differentiation is to represent, in some fashion, the effect of a singularity at P_2 on a test-particle at P_1 . It is therefore to become infinite when $P_1 \rightarrow P_2$ at the same time that $t_1 \rightarrow t_2$.

Second-order covariant differentiation of ϕ_{21} yields the covariant tensor

$$\frac{\partial^2 \phi_{21}}{\partial x_1^\alpha \partial x_2^\beta}.$$

We shall choose our signs so that the anti-symmetrical part of this tensor is defined by relations similar to

$$T_{23} = \frac{\partial^2 \phi_{21}}{\partial y_1 \partial z_2} - \frac{\partial^2 \phi_{21}}{\partial z_1 \partial y_2}, \quad \dots \quad (44)$$

$$T_{14} = \frac{1}{c} \left[\frac{\partial^2 \phi_{21}}{\partial x_1 \partial t_2} - \frac{\partial^2 \phi_{21}}{\partial x_2 \partial t_1} \right], \quad \dots \quad (44')$$

and, of course,

$$T_{23} = -T_{32}, \quad T_{14} = -T_{41},$$

$$T_{11} = T_{22} = T_{33} = 0.$$

Here, for convenience, we have written (x_1, y_1, z_1, ct_1) , (x_2, y_2, z_2, ct_2) for $(x_1^1, x_1^2, x_1^3, x_1^4)$, $(x_2^1, x_2^2, x_2^3, x_2^4)$. The components (T_{23}, T_{31}, T_{12}) of this tensor constitute a 3-vector which we shall call \mathbf{H}_1 ; the components (T_{14}, T_{24}, T_{34}) constitute another vector which we shall call \mathbf{E}_1 . We do not thereby presuppose any physically-known properties of magnetic and electric intensities; $\mathbf{H}_1, \mathbf{E}_1$ remain to be *identified* with magnetic and electric intensities *a posteriori*, from their properties. But the perusal of formulæ is made considerably simpler by using a notation which has been hallowed by custom.

50. Temporarily we now omit the suffix 1, and consider the anti-symmetrical tensor T , or 6-vector (\mathbf{H}, \mathbf{E}) as expressing in some way the state of affairs, consequent on the presence of a singularity at (P_2, t_2) ,

moving with velocity \mathbf{V}_2 , at a particle at (\mathbf{P}, t) , moving with velocity \mathbf{V} . Thus the scheme is

$$\begin{pmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ T_{41} & T_{42} & T_{43} & T_{44} \end{pmatrix} = \begin{pmatrix} 0 & H_z & -H_y & E_x \\ -H_z & 0 & H_x & E_y \\ H_y & -H_x & 0 & E_z \\ -E_x & -E_y & -E_z & 0 \end{pmatrix} \quad (45)$$

51. We now form the inner product of this tensor with the velocity 4-vector \mathbf{V}^ν or $(\mathbf{V}/Y_1^{\frac{1}{2}} \ c/Y^{\frac{1}{2}})$, in connexion with the metric

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dx^2 - dy^2 - dz^2 + c^2 dt^2.$$

The contravariant form of this inner product is

$$g^{\alpha\mu} \mathbf{V}^\nu T_{\mu\nu},$$

whose x -component is

$$c \frac{E_x}{Y^{\frac{1}{2}}} + \frac{v H_z - w H_y}{Y^{\frac{1}{2}}},$$

and whose t -component is

$$\frac{u E_x + v E_y + w E_z}{Y^{\frac{1}{2}}}.$$

These components form the 4-vector

$$\frac{1}{Y^{\frac{1}{2}}} (c\mathbf{E} + \mathbf{V} \wedge \mathbf{H}, \mathbf{E} \cdot \mathbf{V}). \quad (46)$$

52. We are aware, through our previous *mathematical* experience of gravitational potentials, that our definition of external force implied by (9) (9') includes an effect of variation of mass with velocity not belonging to the gravitational potential *per se*. In fact we rapidly find, if we try to let the 4-vector (46) represent an external force (\mathbf{F}, F_t) in (9), (9') that we are led to a contradiction; the time-component of the set is in contradiction with the 3-vector component. This implies that we must go cautiously, and we accordingly set now

$$\mathbf{F} = \frac{k}{Y^{\frac{1}{2}}} \left[\mathbf{E} + \frac{\mathbf{V} \wedge \mathbf{H}}{c} \right] + \alpha \frac{\mathbf{V}}{Y^{\frac{1}{2}}}, \quad (47)$$

$$F_t = \frac{k}{Y^{\frac{1}{2}}} \frac{\mathbf{E} \cdot \mathbf{V}}{c} + \alpha \frac{c}{Y^{\frac{1}{2}}}, \quad (47')$$

leaving α as a 4-scalar to be determined, and k as another scalar.

We now insert (47), (47') in the equations of motion (9), (9') and make them self-consistent. The easiest method is to use the identical relation (10'). We find at once

$$\alpha = \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (m \xi^{\frac{1}{2}}). \quad (48)$$

Introducing (47), (47') into the energy-equation (10) we then find

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (mc^2 \xi^{\frac{1}{2}}) = k \frac{Y^{\frac{1}{2}}}{Z} \left[\frac{\mathbf{E} \cdot (\mathbf{V}t - \mathbf{P})}{Y^{\frac{1}{2}}} - \frac{\mathbf{P} \wedge \mathbf{V} \cdot \mathbf{H}}{c Y^{\frac{1}{2}}} \right]. \quad (49)$$

Here the contents of the square bracket can be shown to be a 4-scalar.

53. The scalar k is at our disposal. Different choices of k will correspond to different natures of test-particle. It is largely immaterial to the theory what is the test particle, but we want it eventually to have similar properties to the field originating particle. Subsequent identifications are rendered most easy if we now choose for k the scalar

$$k = \frac{e}{t_0} \frac{Z}{Y^{\frac{1}{2}}},$$

where e is an arbitrary constant. This then yields from (49) the scalar energy-equation

$$\frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (mc^2 \xi^{\frac{1}{2}}) = e \frac{t}{t_0} \frac{1}{Y^{\frac{1}{2}}} \left[\mathbf{E} + \frac{\mathbf{V} \wedge \mathbf{H}}{c} \right] \cdot \left(\mathbf{V} - \frac{\mathbf{P}}{t} \right), \quad (50)$$

as a consequence of the external force

$$\mathbf{F} = \frac{e}{t_0} \frac{Z}{Y^{\frac{1}{2}}} \left[\mathbf{E} + \frac{\mathbf{V} \wedge \mathbf{H}}{c} \right] \frac{1}{Y^{\frac{1}{2}}} + \frac{\mathbf{V}}{Y^{\frac{1}{2}}} \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (m \xi^{\frac{1}{2}}), \quad (51)$$

$$\mathbf{F}_t = \frac{e}{t_0} \frac{Z}{Y^{\frac{1}{2}}} \left[\frac{\mathbf{E} \cdot \mathbf{V}}{c} \right] \frac{1}{Y^{\frac{1}{2}}} + \frac{c}{Y^{\frac{1}{2}}} \frac{1}{Y^{\frac{1}{2}}} \frac{d}{dt} (m \xi^{\frac{1}{2}}). \quad (51')$$

54. Several comments now suggest themselves. First, in gravitation theory we found that the corresponding scalar α was double the value required by (48). This suggests an essential difference between the "force" of gravitation and the force we are at present introducing, in their effects at high velocities. For α governs the term representing the effect of change of velocity on mass.

It is now a consequence of (51) and (51'), taken in combination with the basic equation of motion (9) and (9'), that motion under the new forces is given in t -measures by

$$\frac{m \xi^{\frac{1}{2}}}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(\frac{\mathbf{V}}{Y^{\frac{1}{2}}} \right) = - \frac{m \xi^{\frac{1}{2}}}{X} \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right) + \frac{e}{t_0} \frac{Z}{Y^{\frac{1}{2}}} \left(\mathbf{E} + \frac{\mathbf{V} \wedge \mathbf{H}}{c} \right) \frac{1}{Y^{\frac{1}{2}}}, \quad (52)$$

$$\frac{m \xi^{\frac{1}{2}}}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(\frac{c}{Y^{\frac{1}{2}}} \right) = - \frac{m \xi^{\frac{1}{2}}}{X} \left(ct - c \frac{Z}{Y} \right) + \frac{e}{t_0} \frac{Z}{Y^{\frac{1}{2}}} \left(\frac{\mathbf{E} \cdot \mathbf{V}}{c} \right) \frac{1}{Y^{\frac{1}{2}}}. \quad (52')$$

55. The second of these two equations might be mistaken for an energy equation. But in our present t -dynamics, energy is a 4-scalar, not the fourth component of a 4-vector, and the true energy equation is the 4-scalar equation (50). This has now an immediate physical interpretation. It represents the rate of increase of kinetic energy, $mc^2 \xi^{\frac{1}{2}}$, as (apart from a factor differing from unity only at high velocities or big distances) the work done by a 3-force

$$\mathbf{E} + \frac{\mathbf{V} \wedge \mathbf{H}}{c}. \quad (51'') \quad (51''')$$

in displacing the particle past its cosmic environment with relative velocity $\mathbf{V}-\mathbf{P}/t$. Thus guided by logical, not physical, ideas, we have been led to associate with the 6-vector (\mathbf{H}, \mathbf{E}) the mechanical force (51). This, of course, we recognize as the pondero-motive force of Larmor and Lorentz.

56. Thirdly, our equations should be valid for unrestricted velocities. The factor $Y^{-\frac{1}{2}}$ in (50), yielding the particular forms of last term occurring in (52) and (52'), proves to be of prime importance when we discuss the high-speed motion of two singularities close to one another.

57. Lastly, our equations contain unexpected secular factors involving the absolute epoch t .

58. The hard work of making a start with the formulation of new forces to represent electromagnetism is now done. We have not yet formulated an expression for the super-potential ϕ_{21} , but we have created, "out of the blue" as it were, a scaffolding out of which to build it. This scaffolding is essentially *dynamical* in nature: that is to say, we have begun with particles in motion, not, as in classical theory, with particles at rest.

59. In the dynamics we constructed in the last paper, we arbitrarily selected a certain invariant, $mc^2\xi^{\frac{1}{2}}$, and subsequently identified it as the counterpart, in the t -dynamics, of energy. Likewise we now seek to construct, out of equations of the type (52) and (52'), something we can call the additional energy due to the presence of the forces defined through \mathbf{H}, \mathbf{E} . We now restore our suffix 1 in equation (50), the energy-equation consequent on (52) and (52'), and write it

$$\frac{d}{dt_1}(mc^2\xi_1^{\frac{1}{2}}) = \frac{e_1}{t_0} \cdot \left(\mathbf{E}_1 + \frac{\mathbf{V}_1 \wedge \mathbf{H}_1}{c} \right) \cdot (\mathbf{V}_1 t_1 - \mathbf{P}_1). \quad (53)$$

In this we substitute expressions for \mathbf{H}_1 and \mathbf{E}_1 in terms of the so far undetermined super-potential ϕ_{21} , according to the relations

$$(\mathbf{H}_1)_x = \frac{\partial^2 \phi_{21}}{\partial y_1 \partial z_2} - \frac{\partial^2 \phi_{21}}{\partial z_1 \partial y_2}, \quad (54)$$

$$(\mathbf{E}_1)_x = \frac{1}{c} \left[\frac{\partial^2 \phi_{21}}{\partial x_1 \partial t_2} - \frac{\partial^2 \phi_{21}}{\partial x_2 \partial t_1} \right]. \quad (55)$$

We form a similar energy equation for the particle m_2 , at (P_2, t_2) in the presence of the field (H_2, E_2) at (P_2, t_2) due to particle m_1 at (P_1, t_1) . We associate with m_2 another constant e_2 , analogous to the e_1 corresponding to m_1 .

60. The algebraic details of the next steps have been given in detail elsewhere. Here I make them a little more general. By multiplying (53) by dt_1/dt and its counterpart by dt_2/dt and adding we find the relation

$$\frac{d}{dt}[m_1 c^2 \xi_1^{\frac{1}{2}} + m_2 c^2 \xi_2^{\frac{1}{2}} + \Phi] = \dot{\mathbf{V}}_1 \cdot \frac{\partial \Phi}{\partial \mathbf{V}_1} \frac{dt_1}{dt} + \dot{\mathbf{V}}_2 \cdot \frac{\partial \Phi}{\partial \mathbf{V}_2} \frac{dt_2}{dt}, \quad (56)$$

provided the superpotentials ϕ_{21} , ϕ_{12} have been chosen to satisfy

$$\phi_{21} + L_{11}\phi_{21} = 0, \quad (57)$$

$$\phi_{12} + L_{22}\phi_{12} = 0, \quad (57')$$

$$-\frac{e_1}{ct_0}L_{12}\phi_{21} = -\frac{e_2}{ct_0}L_{21}\phi_{12} = \Phi. \quad (58)$$

Here L_{11} , L_{22} , L_{21} , L_{12} denote linear differential operators defined by

$$L_{\mu\nu} = \epsilon_{\mu} \frac{\partial}{\partial t_{\nu}} + P_{\mu} \cdot \frac{\partial}{\partial P_{\nu}} \quad (\mu, \nu = 1, 2). \quad (59)$$

61. It must be explained here that we are *constructing* an electrodynamics in the same way as a pure mathematician constructs a geometry. What we are constructing is largely at our disposal. Our aim is to construct something containing features which can subsequently be recognized as the features of empirical electrodynamics. With this object in view we have *imposed* conditions (57), (57') and (58) in order to force our energy equation into the form (56). This form is readily recognized as expressing the rate of increase of the sum of the kinetic energies $m_1 c^2 \xi_1^{\frac{1}{2}}$ and $m_2 c^2 \xi_2^{\frac{1}{2}}$ of the particles together with a function Φ , as equal to an expression linear in the accelerations \dot{V}_1 and \dot{V}_2 of the two particles. Scrutiny of the algebraic details, here omitted, shows that (57) and (57') are imposed to remove terms independent of the accelerations that would be troublesome to interpret physically in the sequel; whilst (58) is imposed to make the left-hand side of (56) the complete t -derivative of a function of the 14 variables concerned, namely the two sets of co-ordinates (x_1, y_1, z_1, t_1) (x_2, y_2, z_2, t_2) and the six velocity component (u_1, v_1, u_1) (u_2, v_2, u_2) .

62. A remark may be made here about the new variable t introduced in deriving (56). This introduction of t is different from the mode of introduction in the original paper in Proc. Roy. Soc. In that paper I adopted a standard of simultaneity $t_1 = t_2 = t$, but the present method, leading to (56), does not require us to specify our convention as to the simultaneity between the epochs at P_1 and P_2 (namely t_1 and t_2) and the epoch t at which observer O reckons the energy. *Some* variable t *must* be introduced before the rate of change of the *sum* of two distinct kinetic energies of two distinct particles is considered, still more when there is an additional term Φ .

63. We shall interpret Φ as the joint energy of the two particles arising from the constants e_1 and e_2 characteristic of them. We shall see later that e_1 and e_2 may be interpreted as the *charges* on the particles, and that Φ will reduce to the ordinary electrostatic energy of the pair of charges in one another's presence. So far we have, however, not given an explicit form for Φ , since neither of the superpotentials ϕ_{21} or ϕ_{12} has yet been fixed. But whatever form we attach to Φ , (56) will be valid when conditions (57), (57') and (58) have been satisfied. Relation

(56), with the interpretation we propose to give to Φ , states that the total energy of the system decreases at the rate

$$\frac{dR}{dt} = - \left[\dot{\mathbf{V}}_1 \cdot \frac{\partial \Phi}{\partial \mathbf{V}_1} \frac{dt_1}{dt} + \dot{\mathbf{V}}_2 \cdot \frac{\partial \Phi}{\partial \mathbf{V}_2} \frac{dt_2}{dt} \right] \dots \dots \dots (60)$$

Since this energy disappears from the energy-sum as computed, we shall call this the *rate of radiation* from the system of two charges. We notice that it is linear in the accelerations. Thus if our interpretation of Φ can be sustained, our theory will provide for absorption as well as emission of radiation, unlike the classical theory where an accelerated charge always *radiates*. It will also provide for cases where systems of accelerated charges suffer no loss by radiation, for the sum of scalar products in (60) can vanish without the accelerations vanishing.

64. We must now consider whether conditions (57), (57') and (58) suffice to determine the super-potentials ϕ_{21} , ϕ_{12} explicitly. The meaning of (57) and (57') is at once clear. Since L_{11} is the Eulerian operator

$$L_{11} = t_1 \frac{\partial}{\partial t_1} + \mathbf{P}_1 \cdot \frac{\partial}{\partial \mathbf{P}_1},$$

(57) means that ϕ_{21} must be homogeneous and of degree -1 in the variables \mathbf{P}_1 , t_1 ; and, similarly, (57') means that ϕ_{12} must be homogeneous and of degree -1 in the variables \mathbf{P}_2 , t_2 .

65. Consider now (58). We are looking out for something that may represent a singularity when the charged particles \mathbf{P}_1 , \mathbf{P}_2 coincide at the same epoch. This therefore suggests as an elementary solution a denominator

$$\frac{1}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}}.$$

But this would not of itself serve as a ϕ_{21} , although it is of dimensions -1 in t_1 , \mathbf{P}_1 , for we find that

$$L_{12} X_{12} = X_1, \quad L_{12} X_1 = 0, \quad L_{12} X_2 = 2X_{12},$$

so that

$$L_{12} \frac{1}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} \equiv 0, \quad \dots \dots \dots (61)$$

and accordingly Φ , which we wish to represent electromagnetic energy, would be zero identically. The elementary function $(X_{12}^2 - X_1 X_2)^{-\frac{1}{2}}$ therefore requires to be multiplied by some scalar function of the variables t_2 , \mathbf{P}_2 and possibly \mathbf{V}_1 and \mathbf{V}_2 , before it can stand for a ϕ_{21} .

Put then tentatively,

$$\phi_{21} = e_2 \frac{\psi_{21}}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}}, \quad \phi_{12} = e_1 \frac{\psi_{12}}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}},$$

where ψ_{21} is some scalar function of the variables t_2 , \mathbf{P}_2 possibly involving \mathbf{V}_1 and \mathbf{V}_2 , and, similarly, ψ_{12} is the corresponding function of t_1 , \mathbf{P}_1 , possibly involving \mathbf{V}_2 and \mathbf{V}_1 . In virtue of (61), condition (58) now requires

$$L_{12} \psi_{21} = L_{21} \psi_{12} \dots \dots \dots (62)$$

The function ψ_{21} must be a function of the scalar products of the three vectors (\mathbf{P}_2, ct_2) , $(\mathbf{V}_1/Y_1^{\frac{1}{2}}, c/Y_1^{\frac{1}{2}})$, $(\mathbf{V}_2/Y_2^{\frac{1}{2}}, c/Y_2^{\frac{1}{2}})$. Write for simplicity

$$Z_{12}=t_1-\mathbf{P}_1 \cdot \mathbf{V}_2/c^2, \quad Z_{21}=t_2-\mathbf{P}_2 \cdot \mathbf{V}_1/c^2.$$

One possible simple solution of (62) at once suggests itself, namely,

$$\psi_{21}=X_2, \quad \psi_{12}=X_1.$$

For

$$L_{12}X_2=2X_{12}=L_{21}X_1.$$

We should then have, by (58),

$$\Phi \propto \frac{X_{12}}{(X_{12}^2-X_1X_2)^{\frac{1}{2}}}.$$

But this is just our old friend χ , the *gravitational* potential energy of two particles, as given by (23). Moreover, such a Φ would not involve V_1 or V_2 , and so we should get $dR/dt \equiv 0$, and by (56) energy would be conserved: there would be no place for radiation phenomena. It is satisfactory, indeed, that our rational gravitational theory prescribes that there is no radiation of gravitational energy. But it is clear that to get something beyond gravitation, we need a solution of (62) involving the velocities of the particles.

The following identities are evident:

$$\begin{aligned} L_{12}Z_2 &= Z_{12}, & L_{12}Z_{21} &= Z_1, \\ L_{21}Z_2 &= Z_{21}, & L_{21}Z_{12} &= Z_2. \end{aligned}$$

It follows that

$$L_{12}(Z_2Z_{21})=Z_{12}Z_{21}+Z_1Z_2=L_{21}(Z_1Z_{12}).$$

The symmetry of this function in the suffixes 1, 2 affords us a possible Φ , by (58), and hence ϕ_{12} and ϕ_{21} . We shall now have a solution of (57), (57') and (58) if we choose

$$\phi_{21} = -\frac{1}{2}e_2 \frac{Z_2Z_{21}}{(X_{12}^2-X_1X_2)^{\frac{1}{2}}Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}}, \quad \dots \quad (63)$$

$$\phi_{12} = -\frac{1}{2}e_1 \frac{Z_1Z_{12}}{(X_{12}^2-X_1X_2)^{\frac{1}{2}}Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}}, \quad \dots \quad (64)$$

$$\text{so that} \quad \Phi = \frac{1}{2} \frac{e_1e_2}{ct_0} \frac{Z_{12}Z_{21}+Z_1Z_2}{(X_{12}^2-X_1X_2)^{\frac{1}{2}}Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}}. \quad \dots \quad (65)$$

For the expression (63) is a 4-scalar, of degree -1 in the variables P_1, t_1 . We could doubtless choose more complicated types of numerators at the cost of having higher powers of $(X_{12}^2-X_1X_2)^{\frac{1}{2}}$ in the denominator. These would correspond to singularities of higher order. But (63), (64), (65) seem to be the simplest type of solution it is possible to choose. The signs have been chosen to make Φ positive when e_1e_2 is positive; and the coefficient $\frac{1}{2}$ is chosen to make Φ subsequently identifiable with energy. The result of all these considerations is that (56) holds good and that ϕ_{21} , and accordingly the values of \mathbf{H} , \mathbf{E} at (\mathbf{P}_1, t_1) , have been determined explicitly. The question now arises: can these values of

E, **H** be recognized as corresponding to the electric and magnetic intensities, arising from a "charge" e_2 at the point P_2 at time t_2 , at the point P_1 at time t_1 ?

66. At first sight the outlook is unpromising. By (44) and (44'), **H**, **E** have to be found by differentiating out ϕ_{21} according to the formulæ

$$(\mathbf{H}_1)_x = \frac{\partial^2 \phi_{21}}{\partial y_1 \partial z_2} - \frac{\partial^2 \phi_{21}}{\partial z_1 \partial y_2}, \quad \dots \dots \dots (66)$$

$$(\mathbf{E}_1)_x = \frac{1}{c} \left[\frac{\partial^2 \phi_{21}}{\partial x_1 \partial t_2} - \frac{\partial^2 \phi_{21}}{\partial x_2 \partial t_1} \right] \dots \dots \dots (66')$$

Since $(X_{12}^2 - X_1 X_2)^{-\frac{1}{2}}$ plays the part of $1/r$, the double partial differentiation of the product of this with a scalar numerator would normally yield terms in $1/r$, $1/r^2$, $1/r^3$. But now a miracle occurs which we did not arrange for. The prospective terms in $1/r^3$ fail to appear on account of the unexpected identities

$$\left(\frac{\partial^2}{\partial y_1 \partial z_2} - \frac{\partial^2}{\partial z_1 \partial y_2} \right) \frac{1}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} = 0, \quad \dots \dots \dots (67)$$

$$\left(\frac{\partial^2}{\partial x_1 \partial t_2} - \frac{\partial^2}{\partial x_2 \partial t_1} \right) \frac{1}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} = 0, \quad \dots \dots \dots (67')$$

whilst the terms in $1/r$ fail to appear on account of the identities

$$\left(\frac{\partial^2}{\partial y_1 \partial z_2} - \frac{\partial^2}{\partial z_1 \partial y_2} \right) Z_2 Z_{21} = 0, \quad \dots \dots \dots (68)$$

$$\left(\frac{\partial^2}{\partial x_1 \partial t_2} - \frac{\partial^2}{\partial x_2 \partial t_1} \right) Z_2 Z_{21} = 0. \quad \dots \dots \dots (68')$$

We are left only with a term in $(X_{12}^2 - X_1 X_2)^{-3/2}$, and we find in fact

$$\mathbf{H}_1 = \frac{e_2}{c^4 Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}}} \frac{\frac{1}{2}(Z_{21} \mathbf{V}_2 + Z_2 \mathbf{V}_1) \wedge (X_2 \mathbf{P}_1 - X_{12} \mathbf{P}_2)}{(X_{12}^2 - X_1 X_2)^{3/2}}, \quad \dots \dots \dots (69)$$

$$\mathbf{E}_1 = \frac{e_2}{c^3 Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}}} \frac{(\mathbf{P}_1 X_2 - \mathbf{P}_2 X_1) \frac{1}{2}(Z_2 + Z_{21}) - (t_1 X_2 - t_2 X_{12}) \frac{1}{2}(\mathbf{V}_1 Z_2 + \mathbf{V}_2 Z_{21})}{(X_{12}^2 - X_1 X_2)^{3/2}}. \quad \dots \dots \dots (69')$$

To see the meaning of these apparently complicated expressions, take the observer **O** to be at P_2 , the source. Then $\mathbf{P}_2 = 0$, $X_2 = t_2^2$, $X_{12} = t_1 t_2$, $Z_{21} = t_2$, $Z_2 = t_2$, and $(X_{12}^2 - X_1 X_2)^{-3/2} = c^3/t_2^3 |\mathbf{P}_1|^3$. We find

$$\mathbf{H}_1 = \frac{e_2}{c} \frac{1}{Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}}} \frac{1}{2} (\mathbf{V}_1 + \mathbf{V}_2) \wedge \frac{\mathbf{P}_1}{|\mathbf{P}_1|^3}, \quad \dots \dots \dots (70)$$

$$\mathbf{E}_1 = e_2 \frac{1}{Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}}} \frac{\mathbf{P}_1}{|\mathbf{P}_1|^3} \dots \dots \dots (70')$$

These are exact inverse square formulæ as regards their distance factors. Moreover, when \mathbf{V}_1 , \mathbf{V}_2 are small, \mathbf{E}_1 as given by (70') reduces to the Coulomb electrostatic formula. This identifies e_2 as charge, and \mathbf{E}_2 as electric intensity.

67. But several features of (70') are of great interest. In the first place, in order that the effect of e_2 (at P_2) on a test charge at P_1 shall be accurately given by an inverse square distance law, the distance $|\mathbf{P}_1 - \mathbf{P}_2|$ must be measured by the fundamental observer O at the second-charge P_2 . Different observers O will assign different values to this distance; that value must be taken which is measured by the observer O at P_2 itself. We see that the observer O is always explicitly in our picture of what we can now call the electric field.

68. In the second place, the value of the distance $|\mathbf{P}_1|$ is the value corresponding to the epoch t_1 at which the field is required at P_1 ; but it is then, when O is at P_2 , explicitly independent of the epoch t_2 at P_2 . Thus, although the epoch t_2 at the source-event has been an indispensable element in our mathematical formulation, it is not necessary in the end to specify the value of t_2 . This does not mean that we could have managed without introducing t_2 , for t_2 disappears only when O is chosen at P_2 . For any other O , the value of $|\mathbf{P}_1 - \mathbf{P}_2|$ will depend on t_2 , and t_2 , as well as t_1 , occurs in a complicated way in formulæ (69). We shall see later the significance of this freedom from mention of t_2 when we come to consider the question of "radiation from an accelerated electron."

69. In the third place, (70') is significantly modified, as compared with the simple Coulomb electrostatic inverse square law, by the occurrence of denominators $Y_1^{-\frac{1}{2}}$, $Y_2^{-\frac{1}{2}}$, whose effect is to increase the intensity when either the source-charge or the test-charge have speeds comparable with c . This is a new feature, not disclosed by electrostatics. That the feature is necessary can be shown by the 6-vector nature of \mathbf{H}_1 , \mathbf{E}_1 . We must point out, however, that it has not been arbitrarily introduced; ϕ_{21} could only be a scalar if it contained factors $Y_1^{\frac{1}{2}}$, $Y_2^{\frac{1}{2}}$ in its denominator.

70. Lastly, (69) represents purely the effect of a *point-charge*, a singularity. Its special form (70') shows no trace of any departure from the inverse square law of distance, however small $|\mathbf{P}_1|$ may be; we have assigned no *structure* to our point singularity. Singularities containing structure, such as a singularity representing a doublet, could be obtained by appropriate differentiation of formulæ already obtained, but (69') represents the \mathbf{E}_1 derived from the simplest form of ϕ_{21} we could choose. It will remain for us to show how a pure point-singularity comes to behave as though it possesses a definite radius. We postpone this for the present.

71. The magnetic intensity, represented by (69), reduces to (70) for O at P_2 . This formula is generally akin to the formula for the magnetic effect of a charge in motion usually attributed to Biot and Savart, to which it reduces for small velocities when $\mathbf{V}_1 = \mathbf{V}_2$. Its principal feature of interest is that whereas in Biot and Savart's form of this magnetic inverse square law we find mention only of \mathbf{V}_2 , the velocity of the source, in our formula (70) we find \mathbf{V}_2 replaced by $\frac{1}{2}(\mathbf{V}_2 + \mathbf{V}_1)$. This means that

the test-charge itself contributes by its motion to the apparent magnetic field in which it finds itself, and that truly speaking a magnetic field only has a definite value, as measured by a test-charge, when the velocity of the test-charge is specified. The velocity $\frac{1}{2}(\mathbf{V}_2 + \mathbf{V}_1)$ may be written in the form

$$\mathbf{V}_2 + \frac{1}{2}(\mathbf{V}_1 - \mathbf{V}_2), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (71)$$

and so we may say that in addition to the Biot and Savart component there is another component, due to the relative velocity of source-charge and rest-charge, and numerically equal to one-half of the field that would be calculated by a crude application of the Biot and Savart law to this relative velocity. The effect has been known for some time—it is due to L. H. Thomas—and its existence is in accordance with elementary relativity. A more complete discussion of the factor one-half which occurs in (71) will be found in my Royal Society paper already cited. Here I shall in due course apply the new formula (70) to the orbital motion of an electron round a nucleus. We note that (70), like (70'), is the magnetic consequence of a structureless singularity.

72. Both (70) and (70') have originated in a rigidly constructed dynamics which has been arranged to hold good for all velocities, however near c . The proper field for testing the validity of the forms (70) and (70') is therefore in the domain of high velocities.

73. We have so far confined attention to two charges in one another's presence, of which one was the source-charge e_2 , the other the test-charge exploring the field. If we wish to discuss the "field" due to more than one source charge, we do so by simple addition of superpotentials and intensities, but we must always be careful to mention both observer O and test-charge e_1 at \mathbf{P}_1 , t_1 . We have to mention the velocity \mathbf{V}_1 of e_1 , as well, before $(\mathbf{H}_1, \mathbf{E}_1)$ the field at \mathbf{P}_1 at t_1 , becomes definite. It might have been feared that since \mathbf{P}_1 and t_1 are so intimately interwoven with the other epochs and position-vectors of the other charges, difficulty would occur in stating the analogues of Maxwell's equations in simple form. Actually, provided differentiation of the field at \mathbf{P}_1 , t_1 is taken with respect to the variables \mathbf{P}_1 , t_1 themselves, \mathbf{V}_1 being kept constant, equations similar to Maxwell's are satisfied whatever is \mathbf{V}_1 , without specification of the vector \mathbf{V}_1 , and without bringing in any relations between the different epochs t_s at charges e_s , or relating them to t_1 . This is very remarkable: it shows that the satisfaction of Maxwell's equations, as such, does not depend on the phenomenon of radiation, which would involve setting relations of retardation between the epochs at the source-charges and the epoch at the test-charge.

74. For completeness of this series of papers, it may be worth while to put on record the various relations analogous to Maxwell's which are satisfied by our $(\mathbf{H}_1, \mathbf{E}_1)$. Before we do so, we must emphasize the difference between the rôles of Maxwell's equations in Maxwell's theory

and those of their analogues in our theory. In Maxwell's theory, the fundamental equations valid in free space, namely,

$$\operatorname{div} \mathbf{E} = 0, \quad \operatorname{div} \mathbf{H} = 0, \quad . \quad . \quad . \quad (72) \quad (72')$$

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}, \quad \operatorname{curl} \mathbf{H} = +\frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}, \quad . \quad . \quad (73) \quad (73')$$

are taken as basis; and in Lorentz's theory of electrons the properties of the electron are derived from these, suitably modified for the presence of volume charge of density ρ . In our present theory, the field is *derived* from the singularities, and the equations we are about to state are not like Maxwell's, *conditions* imposed on the field but *identities* satisfied by the field.

75. We define our operators *div*, *curl* by

$$\operatorname{div} \mathbf{H}_1 \equiv (\operatorname{div})_1 \mathbf{H}_1 \equiv \frac{\partial (\mathbf{H}_1)_x}{\partial x_1} + \frac{\partial (\mathbf{H}_1)_y}{\partial y_1} + \frac{\partial (\mathbf{H}_1)_z}{\partial z_1}, \quad . \quad . \quad . \quad (74)$$

$$\operatorname{curl} \mathbf{H}_1 \equiv (\operatorname{curl})_1 \mathbf{H}_1 \equiv \frac{\partial (\mathbf{H}_1)_z}{\partial y_1} - \frac{\partial (\mathbf{H}_1)_y}{\partial z_1}, \quad . \quad . \quad . \quad (75)$$

etc. The field \mathbf{H}_1 , \mathbf{E}_1 is defined at (\mathbf{P}_1, t_1) by a test-charge moving there with velocity \mathbf{V}_1 by

$$(\mathbf{H}_1)_x = \sum_{s=2}^n \left[\frac{\partial^2 \phi_{s1}}{\partial y_1 \partial z_s} - \frac{\partial^2 \phi_{s1}}{\partial z_1 \partial y_s} \right], \quad . \quad . \quad . \quad (76)$$

$$(\mathbf{E}_1)_x = \sum_{s=2}^n \left[\frac{\partial^2 \phi_{s1}}{\partial x_1 \partial t_s} - \frac{\partial^2 \phi_{s1}}{\partial t_1 \partial x_s} \right], \quad . \quad . \quad . \quad (77)$$

where the superpotentials ϕ_{s1} are defined through the source-charges e_s at (\mathbf{P}_s, t_s) moving with velocity \mathbf{V}_s , by

$$\phi_{s1} = -\frac{\frac{1}{2} e_s}{Y_1^{\frac{1}{2}} Y_s^{\frac{1}{2}}} \frac{Z_s Z_{s1}}{(X_{1s}^2 - X_1 X_s)^{\frac{1}{2}}}, \quad . \quad . \quad . \quad (78)$$

with

$$\begin{aligned} X_1 &= t_1^2 - \mathbf{P}_1^2/c^2, & X_s &= t_s^2 - \mathbf{P}_s^2/c^2, \\ X_{1s} &= t_1 t_s - \mathbf{P}_1 \cdot \mathbf{P}_s/c^2, \\ Y_1 &= 1 - \mathbf{V}_1^2/c^2, & Y_s &= 1 - \mathbf{V}_s^2/c^2, \\ Z_s &= t_s - \mathbf{P}_s \cdot \mathbf{V}_s/c^2, & Z_{s1} &= t_s - \mathbf{P}_s \cdot \mathbf{V}_1/c^2. \end{aligned}$$

The electric energy of the system of charge is Φ , given by

$$\Phi = \sum_{r,s} \Phi_{r,s}, \quad . \quad . \quad . \quad (79)$$

extended to every pair of charges e_r , e_s , where

$$\Phi_{r,s} = \frac{\frac{1}{2} e_r e_s}{c t_0} \frac{Z_r Z_s + Z_r Z_s}{(X_{rs}^2 - X_r X_s)^{\frac{1}{2}} Y_r^{\frac{1}{2}} Y_s^{\frac{1}{2}}}, \quad . \quad . \quad . \quad (80)$$

The energy satisfies the mechanical equation

$$\frac{d}{dt} \left[\sum_{s=1}^n m c^2 \xi_s^{\frac{1}{2}} + \Phi \right] = \sum_{s=1}^n \dot{\mathbf{V}}_s \cdot \frac{\partial \Phi}{\partial \mathbf{V}_s} \frac{dt_s}{dt}. \quad . \quad . \quad . \quad (81)$$

The equation of motion of each charge is of the form (52) above, with each symbol appropriately suffixed.

76. It can then be shown, from these specific definitions of \mathbf{H}_1 , \mathbf{E}_1 , that

$$\operatorname{div} \mathbf{H}_1 = 0, \quad \dots \dots \dots (82)$$

$$\operatorname{curl} \mathbf{E}_1 = -\frac{1}{c} \frac{\partial \mathbf{H}_1}{\partial t_1}, \quad \dots \dots \dots (83)$$

but that

$$\operatorname{div} \mathbf{E}_1 = -\frac{1}{c} \frac{\partial a_1}{\partial t_1}, \quad \dots \dots \dots (84)$$

$$\operatorname{curl} \mathbf{H}_1 - \frac{1}{c} \frac{\partial \mathbf{E}_1}{\partial t_1} = \frac{\partial \mathbf{a}_1}{\partial \mathbf{P}_1}, \quad \dots \dots \dots (85)$$

where a_1 is a very small scalar given by

$$a_1 = \sum_{s=2}^n \square_1 \cdot \square_s \phi_{s1}, \quad \dots \dots \dots (86)$$

and $\square_1 \cdot \square_s$ stands for the scalar operator

$$\square_1 \cdot \square_s = \frac{\partial}{\partial \mathbf{P}_1} \cdot \frac{\partial}{\partial \mathbf{P}_s} - \frac{1}{c^2} \frac{\partial}{\partial t_1} \frac{\partial}{\partial t_s}.$$

It is found that whilst

$$\square_1^2 \phi_{s1} \equiv 0, \quad \dots \dots \dots (87)$$

we have

$$\square_1 \cdot \square_s \phi_{s1} \equiv \neq 0.$$

Thus whilst (82), (83), agree with Maxwell's equations (72'), (73), the further identities (84) (85) differ slightly from Maxwell's (72) (73'). This might perhaps be expected. Maxwell's equation (73) expresses Faraday's law of electromagnetic induction, and so we should have had a serious disagreement with experience had not our (83) been fulfilled. Similarly, Maxwell's (72') states the non-existence of magnetic poles, which is an essential part of our theory. But Maxwell's (73') is hypothesis, the hypothesis of the displacement current, introduced for certain never very convincing physical reasons, and we are not surprised that when we avoid making this hypothesis we arrive at the slightly different equation (85). Likewise when we dispense with the possibility of a volume distribution of electric charge (save macroscopically), we find (72) replaced by (84). We can eliminate the small scalar a_1 from (84) and (85) by cross differentiation, obtaining the six equations

$$\operatorname{grad} \operatorname{div} \mathbf{E}_1 + \frac{\partial}{c \partial t_1} \left[\operatorname{curl} \mathbf{H}_1 - \frac{1}{c} \frac{\partial \mathbf{E}_1}{\partial t_1} \right] \equiv 0, \quad \dots \dots (88)$$

$$\operatorname{curl} \left[\operatorname{curl} \mathbf{H}_1 - \frac{1}{c} \frac{\partial \mathbf{E}_1}{\partial t_1} \right] \equiv 0, \quad \dots \dots \dots (89)$$

which now contain no mention of the source-charges e_s and so may legitimately be considered field identities.

77. It is easy to show, either from (87) or the set (82), (83), (88), (89), that \mathbf{E}_1 and \mathbf{H}_1 satisfy the wave-equations

$$\square_1^2 \mathbf{E}_1 \equiv 0, \quad \square_1^2 \mathbf{H}_1 \equiv 0. \quad \dots \dots \dots (90) \quad (90')$$

Further, \mathbf{H}_1 , \mathbf{E}_1 , can be expressed in terms of a vector potential (\mathbf{A}_1 , A_1) defined by

$$\mathbf{A}_1 = \sum_{s=2}^n \frac{\partial \phi_{s1}}{\partial \mathbf{P}_s}, \quad A_1 = \sum_{s=2}^n \frac{\partial \phi_{s1}}{c \partial t_s}, \quad \dots \quad (91) \quad (91')$$

for then

$$\mathbf{E}_1 = \frac{\partial \mathbf{A}_1}{\partial t_1} - \frac{\partial \mathbf{A}_1}{c \partial t_1} = \text{grad } A_1 - \frac{1}{c} \frac{\partial \mathbf{A}_1}{\partial t_1}, \quad \dots \quad (92)$$

and

$$\mathbf{H}_1 = \text{curl } \mathbf{A}_1, \quad \dots \quad (92')$$

These suggest that for *stationary* charges, A_1 should be the negative of the electrostatic potential. That this is so follows by taking the origin at the test-charge, so that $\mathbf{P}_1 = 0$, when we find

$$\sum_{s=2}^n \Phi_{s,1} = - \sum_{s=2}^n \frac{e_1}{ct_0} L_{1s} \phi_{s1} = - \frac{e_1 t_1}{ct_0} \sum_{s=2}^n \frac{\partial \phi_{s1}}{\partial t_s} = - \frac{t_1}{t_0} e_1 A_1,$$

whilst by actual evaluation, for $\mathbf{V}_1 = 0$, $\mathbf{V}_s = 0$,

$$\sum_{s=2}^n \Phi_{s,1} = \sum_{s=2}^n \frac{e_1 e_s t_s}{|\mathbf{P}_s| t_0},$$

so that at the epochs $t_s = t_1 = t_0$,

$$A_1 = - \sum_{s=2}^n \frac{e_s}{|\mathbf{P}_s|}, \quad \dots \quad (93)$$

as in electrostatics.

78. It follows that the usual electromagnetic theory of the propagation of light will follow from the present treatment. The new term a_1 we have found as modifying the simple scheme of Maxwell's equations in free space is found to be numerically very small. As mentioned before, the set of equations (82), (83), (88), (89), which replaces Maxwell's set, holds good without our having had to specify \mathbf{V}_1 , the velocity of the test charge, or to specify relations between the t_s 's, the epochs at the source charges, and t_1 , the epoch at the test-charge. Thus the equations do not *assume* wave-propagation, but imply it as a consequence in all contexts where, to evaluate formulæ, it is necessary to state relations between epochs.

Energy is not calculated as located in the field, but is a joint affair of pairs of charges, as measured by Φ . The fact that the sum of Φ and the mechanical energies $mc^2 \dot{\xi}_s^{\frac{1}{2}}$ of the charged particles does not necessarily remain constant when accelerations are present, permits energy to be transformed into radiation in non-stationary fields, but we do not get the absurd consequence associated with Poynting's theorem in the classical theory, by which radiation is associated with the existence of crossed electrostatic and magnetic fields even when stationary.

Clearly much more work remains to be done to elucidate the consequences of the present theory as regards radiation phenomena, especially the radiation of momentum. As a preliminary investigation of this type, I propose in the next paper to examine the relative motion of two charges, when the field exerted by each on the other is given by the formulæ we have here obtained.

XXIII. *Fluorescence Notation.*

By JACK DE MENT*.

[Received July 1, 1942.]

WITH the recent definition of a new field of chemical science known as *fluorochemistry* ⁽¹⁾, a need has arisen for a notation with which to denote the simple fluorescence of a substance, its exciting wave-length, and the conditions, *e. g.*, solvent, under which the examination was conducted.

As the term fluorochemistry was created to identify a field distinctly apart, although somewhat analogous to photochemistry and other specialized branches of chemistry and/or physics, it seems expedient to adopt a simple notation presenting maximal information. Therefore, it is proposed that the following notation be employed to indicate fluorescence as a property distinct and characteristic of a chemical compound under filtered ultra-violet light, together with the solvent or other treatment which may have been accorded the substance before examination :

$$[A]_{\text{solvent}}^1 = \text{Fluorescence colour.}$$

For riboflavin in water, for example, the notation would read,

$$[A]_{\text{H}_2\text{O}}^{3650} = \text{Bright green.}$$

But in the solid state without a solvent the fluorescence notation for riboflavin would read,

$$[A]_{-}^{3650} = \text{Orange.}$$

The urgent need for a standardized notation which connotes fluorescent response (as seen by the unaided eye and not by instrumental means, *e. g.*, the spectroscope), wave-length of ultra-violet light used for excitation (filtered), and conditions under which the compound was examined, are well known to those active in the field. As fluorescence is being described more and more in terms of colour, especially when fluorochemical analysis is involved, the notation proposed may tend to reduce the ambiguous, and often vague, results which have been so prevalent among investigators who neglect to note the exciting wave-length employed ⁽²⁾.

Literature.

- (1) J. De Ment, 'Science,' xcv, p. 407 (1942).
- (2) J. De Ment, 'Fluorescent Chemicals and their Applications,' New York (1942).

* Communicated by the Author.

XXIV. *Contact Potentials.**To the Editors of the Philosophical Magazine.*

GENTLEMEN,—

IN the first of the series of papers in this Journal by one of us (J. A. C.] on "Contact Potentials," the criticism was made that in the book 'The Metallic State,' the other (W. H. R.] had put forward entirely false conclusions when, in discussing surface layers, the statement was made that "the contacts . . . are no longer of a simple physical nature." Correspondence between us has revealed that this criticism was based on a misunderstanding of the book. The actual sequence of statements in 'The Metallic State' was (1) a repetition of the conventional argument that the contact potential between two metals A and B should be unaffected by the interposition of C; (2) the statement that it was doubtful whether this argument is justified when C consists of oxide films, because the contacts of these are not of a simple physical nature; and (3) the statement that contact potentials are frequently affected by surface conditions. The argument was not carried further, and the book then discussed the unsatisfactory nature of the experimental data. We are in agreement that the book might be criticized for not differentiating between the effects of surface layers at metal-metal, metal-dielectric, and metal-electrolyte contacts, as has been done in the recent series of papers, where it has been pointed out that (1) above is still true for most metal-metal contacts, while the result (3) is caused by films at metal-dielectric contacts. The comment in 'The Metallic State' that oxide films do not involve simple physical contact was, however, not put forward as a conclusion, but was made as a simple statement of experimental fact, and is undoubtedly correct in the majority of cases, if by simple physical contact we mean the kind of contact produced when two clean surfaces are pressed together in a vacuum; from the results in "Contact Potentials" it follows that conclusion (1) above does not require simple physical contacts in this sense of the term, but is much more widely true, and the warning in 'The Metallic State' as to the possible effects of the nature of the contacts does not happen to be necessary in many cases.

The statement that, in 'The Metallic State,' an entirely false conclusion was reached regarding the nature of contacts should therefore be withdrawn.

Yours faithfully,

J. ALAN CHALMERS,

W. HUME-ROTHERY.

XXV. *Notices respecting New Books.*

The Life of Sir J. J. Thomson. By LORD RAYLEIGH. [Pp. x+299.]
(Cambridge University Press, 1942. Price 18s.)

THE presentation of science is often so dogmatic and impersonal that the extent to which the scientific outlook of any age is a reflexion of the intensely personal thoughts and actions of individual scientists is hidden from view. The individual contribution is generally so small, in relation to the whole, that this is inevitable. Indeed, for the assessment of a piece of scientific work, the personal background may be irrelevant. But the answer to the question of how a particular advance was made, whether it be an experimental discovery or a Euclidean theorem, is a personal story. There can be no better corrective to that false conception of science as something divorced from the humanities than the lives of those who have contributed greatly to the scientific heritage on which we build to-day. How great discoveries were made is of particular interest to the professional scientist and the science student, and the theme can be admirably exemplified in the life of J. J. Thomson, who, in the words of Bragg, "more than any other man, was responsible for the fundamental change in outlook which distinguishes the physics of this century from that of the last." Beyond this there is a more general interest. What kind of men were these who, by common consent, have achieved greatness? Their work alone may give the essence of an answer, but for some, like J. J., how much more there is! His long life (1856-1940), too, covered a period of immense change in every sphere, and the story of Thomson and the circle with which he was associated gives a fascinating cross-section of history.

Lord Rayleigh is peculiarly qualified for the task of writing this biography. The family association dates back to before 1884, when J. J. succeeded the biographer's father as Cavendish Professor at Cambridge. Rayleigh was an undergraduate at Trinity, and a research worker at the Cavendish Laboratory during some of J. J.'s greatest days there, and in later years he was a close friend of the Thomson family. Moreover, his own wide knowledge and particular interests in experimental physics enable him to appreciate fully not only the achievement, but also the difficulties, of J. J.'s scientific work.

It is not a formal full-scale biography that Lord Rayleigh has written. The solemn tomes in which the lives of some of our past scientists are set forth, based perhaps on assiduous delving into records and characterized by a meticulous transcription of every detail that could be unearthed, no doubt have their value; but they often conceal rather than reveal. Rayleigh's treatment is more impressionistic. He presents Thomson by the consideration of a series of main themes, rather than by a day by day or even year by year record of the events of his life. He brings together relevant material, letters, extracts from speeches and anecdotes in considering these themes with little detailed regard to chronology, though the order of the themes themselves is chronologically appropriate. The result, as a whole, is a vivid impression of J. J.'s work and personality. But there is much about which many would have liked more—Thomson's theoretical work, and his personal friendships, such as that with Poynting, may be mentioned as examples. Rayleigh was perhaps too modest in his approach. He writes, in the preface, of Thomson's 'Recollections and Reflections' (1936) having skimmed much of the cream off what there is to be told, and of his own inclining to the view, when asked to write a biography, that the subject was almost exhausted. Nothing could be further from the truth, as Rayleigh must have realized in the course of writing this book.

The early years are passed over very rapidly—by page 19, Thomson has been

appointed Cavendish Professor (1884). J. J.'s recollections do give other glimpses of this early period, but it is none the less surprising that the accessible information about his early life seems almost as fragmentary as that about Newton's. There would be much of interest in a fuller account of the development of one who was accepted at the age of twenty-eight as a worthy successor to Maxwell and Rayleigh.

The biography effectively begins with the Cavendish Professorship. The background of J. J.'s scientific life is skilfully portrayed, and admirable accounts are given of the scientific work on the electron and on positive rays with which J. J.'s name will always be associated. The two great series of investigations both had their origin in the experimental study of the discharge of electricity through gases. The first peak was reached about 1897, when the electron may be said to have been "discovered." Experiments were made, that is to say, which demonstrated conclusively the existence as a common constituent of all matter of negatively charged corpuscles much smaller in mass than any atom, and of unique charge and mass. The early work in the development of the methods for the quantitative study of the properties of these electrons, in the elucidation of their rôle in a wide range of phenomena, and in the new attack on the general problem of the constitution of matter which was made possible, was to a remarkable extent carried out by J. J. himself and that band of research workers who gathered round him in the Cavendish Laboratory in the closing years of the last and the opening years of the present century. These days at the Cavendish were afterwards described by J. J. as "the happiest days of my life." In Rayleigh's biography we see him full of enthusiasm and abounding vitality, making the round of the laboratory, a stimulating leader of the research workers with brilliant and sometimes incredible suggestions, the centre of the lively conversation at the informal teas, and showing foreign visitors round the laboratory, when his resonant voice reached its greatest intensity through his apparent belief that so could English be made intelligible to those ignorant of the tongue.

In these days when "electron" is one of the first words in physico-chemical discourse, it is difficult to appreciate fully the degree of theoretical insight and of experimental ingenuity of the pioneers. Rayleigh's treatment of the course of events is particularly good in this connexion for he presents the story as it unfolded itself at the time, with only that desirable minimum of comment in the light of later knowledge which makes for clearness and brevity. The second great period of experimental achievement Rayleigh regards as culminating about 1912, with the development of the beautiful parabola method of investigating positive rays. Rayleigh's account of all this scientific work is simple and straightforward. That he has used no formulæ, "in view of the protest of a well-known literary man, that he could not skip even a formula," does not, however, call for unqualified approval. Physics, after all, does make use of formulæ, whether or not literary men are reading about it, and one of the purposes of formulæ is to make things simpler. The self-denying ordinance, moreover, makes very difficult a proper presentation of one side of J. J.'s scientific genius. It was sometimes felt that J. J. in his later work, with his view of a theory as "a policy and not a creed," did not always follow fairly the rules of the game, and that his speculations were remarkable for ingenuity, and a disregard of difficulties, rather than for revelation. But the aggregate of his solid theoretical work was very considerable, and alone would have ensured him a lasting name among scientists.

The war years saw Thomson taking a leading part on the Board of Invention and Research, the main committee of which met under the chairmanship of Lord Fisher. He enjoyed the wider contacts which this involved, and spent much time on the work, though Rayleigh does not think it was particularly fruitful as regards ultimate results.

In 1915 Thomson accepted nomination as President of the Royal Society.

He held that office until 1920, and carried on the traditions in an admirable way through a difficult period.

A new phase of his life began in 1918, when he was appointed Master of Trinity. He hoped at first that this would not diminish the time he could give to science. After his resignation from the Cavendish Professorship in 1919, and Rutherford's appointment as his successor, Thomson indeed continued as honorary Professor of Physics. There might well have been difficulties in the situation, and it says much for the characters of the two men that the transition in the control of the Laboratory was effected with so little friction. J. J. continued to give lectures at the Cavendish; lectures which many post-war undergraduates and graduates feel they were indeed privileged to attend, whether or not the views propounded fitted into any orthodox frame. J. J. also continued his experimental work for many years, but Trinity took a larger and larger share of his time—not so much because of the routine duties of the Mastership, but because of his enthusiastic concern with every detail of the College activities. His friendly interest in the less intellectual undergraduates, and his keenness about the athletic side of college life sometimes surprised others; the degree to which he endeared himself to the Fellows perhaps surprised himself. The demonstration that a scientist could become a great Master of Trinity is perhaps not the least of J.J.'s achievements.

Rayleigh includes a most interesting chapter in which he brings together J. J.'s views on education, views which were largely put forward in occasional speeches. "The general keynote," Rayleigh says, "is that he was entirely opposed to any kind of pedantry or formalism in education," and it is not surprising that no self-consistent educational programme can be built up from his remarks. But his statements on individual topics—on elementary teaching, lectures and reading, examinations, classical and mathematical education, the educational value of research—are often full of wisdom. In some respects J. J. was a thorough reactionary. He had, for example, little enthusiasm for raising the school leaving age, and his attitude in connexion with the status of women at Cambridge is notorious. In the last chapter reference is made to some of J. J.'s more personal qualities and characteristics—his religious sincerity, his tastes in reading, his love of wild flowers.

There was something Johnsonian about J. J. in his vitality and assurance, and in his apparent disregard of what others might think. There was a suggestion of unapproachability, alarming to the diffident, and yet, Rayleigh says, "he was the most accessible person in the world." There are many other paradoxes. A world figure for his experimental work, he was himself clumsy with apparatus. Aston remarks on his "intuitive ability to comprehend the inner working of intricate apparatus without the trouble of handling it as something verging on the miraculous, the hall mark of a great genius." His ideal arrangement for the research worker was "to pay him for something else and give him enough leisure to do research for the love of it." His conceptions of the appropriate emoluments of laboratory teachers and of scientific workers generally suggest, however, that the payments for "something else" were not envisaged on a very lavish scale, although he could not have been unaware of the advantages of his own freedom from financial cares. There was, however, a sincerity and a forthrightness in all that J. J. did and said which revealed all the inconsistencies which lesser men are at pains to conceal. There could be no better comment on the effect he produced than that of Rayleigh about his Cavendish research students, "We might perhaps laugh at his little peculiarities, but we knew he was a great man, and we all loved him." Rayleigh has given a vivid picture of this great man, and it will give some happy hours to all those who are interested in science and scientists, and in an outstanding personality.

E. C. S.

[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]

XXVI. *The Determination of Velocity of Sound by the Employment of Closed Resonators and the Hot-wire Microphone.*

By W. S. TUCKER, D.Sc., A.R.C.Sc.*

[Received December 30, 1942.]

Summary.

A modified form of the Helmholtz resonator, consisting of two resonating cavities connected by a narrow neck, is employed for measurements of velocity of sound. Sound is introduced by a telephone diaphragm forming the boundary of one enclosure, and both temperature and response of the resonator to sound are indicated by the change in electrical resistance of a hot-wire in the neck of the resonator. Frequency-response curves are obtained from which the true resonance frequency maximum is obtained. Theoretical support is given of a linear relationship between velocity and resonance-frequency for different gases, and this has been confirmed by experiment.

Determinations of velocity of sound for air saturated with water vapour were made and data obtained for calculation of γ . From this the velocity of sound in water vapour at 0° C. was deduced, and the result compared with those values derived by other investigators.

Results were also obtained for the velocity of sound in ether and acetone vapours by a similar process.

§ 1. *Introduction.*

THE Helmholtz resonator within certain limits has a natural frequency of vibration which is directly proportional to the velocity of sound of the air in its orifice. If its dimensions are small compared with the wave-length of the incident sound it gives single frequency response, and the relation between its frequency n and its dimensions can be expressed by the relation

$$n = \frac{c}{2\pi} \sqrt{\frac{k}{q}},$$

where c is the velocity of sound, k a function only of the dimensions of the neck and q the volume. That n is in linear relation with c to a good degree of accuracy can be proved experimentally by a method described by E. T. Paris and the author, employing the hot-wire microphone as a detector of motions of air in the neck †.

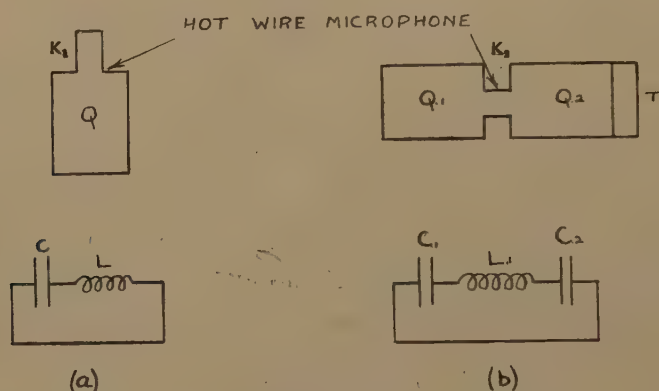
* Communicated by the Author.

† Tucker and Paris, Phil. Trans. Series A, cccxi, pp. 389-430.

Further experiments have also shown that the frequency at resonance is not appreciably affected by the variation of intensity of the incoming sound within wide limits. The Helmholtz resonator is an acoustical analogue of the simple oscillating electric circuit in which the air in the orifice is in the nature of an inductance while the volume of the resonator takes the place of a capacitance (fig. 1 (a)).

The device here described may be regarded as a development of the Helmholtz resonator. It consists of two such resonators with a common neck as shown in the diagram (fig. 1 (b)). The relative volumes of these constituent resonators may be varied within reasonable limits, but the work here described employed two volumes nearly equal in magnitude. Pursuing the electrical analogy the modified resonator, which will be referred to as the double-Helmholtz resonator, can also be compared with an oscillating circuit in which two condensers, one on either side of the inductance, can be visualized, as in the diagram. As the "con-

Fig. 1.



ductivities" of the orifice k will not be precisely the same owing to variation of the end conditions for the single and double resonators, they may be distinguished as K_1 and K_2 , but the capacitances C_1 and C_2 are expressible in both cases as the volumes Q_1 and Q_2 .

Justification for this interpretation can be provided by experiment. The method employed is to observe the response of these resonating systems to an injected sound of frequency which can be varied continuously, using as detector the hot-wire microphone to which reference has already been made. With a single resonator the sound is provided by a loud-speaker in the neighbourhood of the orifice—with the double-resonator one end of the cavity Q_2 is walled by the diaphragm of a telephone receiver T .

If single resonators fitted with the same orifice have volumes q' and q'' , the frequencies n' and n'' should give $\frac{n'}{n''} = \sqrt{\frac{q''}{q'}}$ and the conductivity K given by $\frac{4\pi^2 q n^2}{c^2} = \text{constant}$ where c is the velocity of sound.

In the comparison q' was 241.4 c.c., and q'' 411.2 c.c. With double resonators having constituent volumes q_1 and q_1' , q_2 and q_2' respectively, the equivalent volumes should be

$$\frac{q_1 q_1'}{q_1 + q_1'} \quad \text{and} \quad \frac{q_2 q_2'}{q_2 + q_2'},$$

which is the analogue of the capacitance of two capacitances in series*.

When compared with the same neck these also should give the same value for K_2 , but rather smaller than K_1 , for the same neck in single resonators. The constituent volumes taken were $q_1 = q_1' = 241.4$ c.c., and $q_2 = 241.4$ c.c., $q_2' = 411.2$ c.c. giving equivalent volumes of 120.7 c.c. and 152.1 c.c. respectively.

The results of these experiments are summarized in Tables I. and II.

TABLE I.

Type of resonator.	Volume or equivalent (c.c.).	Resonance frequency (cyc/s.).	Conductivity K (cm.).	Mean.
Single 1....	241.4	153	0.1883	$K_1 = 0.1895$
Single 2....	411.2	118	0.1908	
Double 3...	120.7	203	0.1657	$K_2 = 0.1631$
Double 4...	152.1	178	0.1605	

The comparison of the two types of resonator is made in Table II.

TABLE II.

Comparison of resonators.	n_1/n_2 .	$\sqrt{\frac{K_1 Q_2}{K_2 Q_1}}$.
1 with 3	0.754	0.762
1 with 4	0.859	0.856
2 with 3	0.581	0.584
2 with 4	0.663	0.655

* Hahnemann and Hecht, *Phys. Zeit.* xxii. p. 357.

When single and double resonators are compared, of volumes such as q' and $\frac{q_2 q_2'}{q_2 + q_2'}$, the ratio of frequencies N_1 and N_2 should be

$$\frac{N_1}{N_2} = \sqrt{\frac{K_1(q_2 + q_2')}{K_2 \cdot q' \cdot q_2 q_2'}}.$$

Within the limits of experimental error therefore it may be accepted that the same expression will be applicable to double as to single Helmholtz resonators and that if one double resonator be employed in a series of velocity determinations the resonance frequency should be proportional to the velocity.

To complete the electrical analogy the inclusion of a resistance in the circuit corresponds with a viscous resistance in the enclosure. If the latter is a good resonating cavity the effect on the resonance frequency should not be serious.

§ 2. Investigation of the resonance-frequency and velocity relation for the Double Helmholtz Resonator.

(1) Description of Apparatus.

A double resonator was constructed from brass tubing of internal diameter 4.9 cm. and about 1 mm. thick. It was made in two equal portions of length internally of 12.8 cm., and the end of one portion was closed with a circular plate (fig. 2, A). The other end was fitted with a brass flange sweated on, and its centre was drilled with a circular hole 0.6 cm. in diameter (B). The second portion of the resonator was made precisely similar except that the closed end was replaced by the cap of a telephone receiver of standard type bolted on to the tube to which a shallow flange had been fitted (C) so as to make an air-tight fit. The central portion of this cap was cut away, leaving only the narrow ebonite flange C to support the telephone diaphragm as normally existing. With the exception of this narrow flange the diaphragm thus made a flat termination corresponding precisely to the rigid base A of the first portion. In this way two equal halves of the resonator were produced.

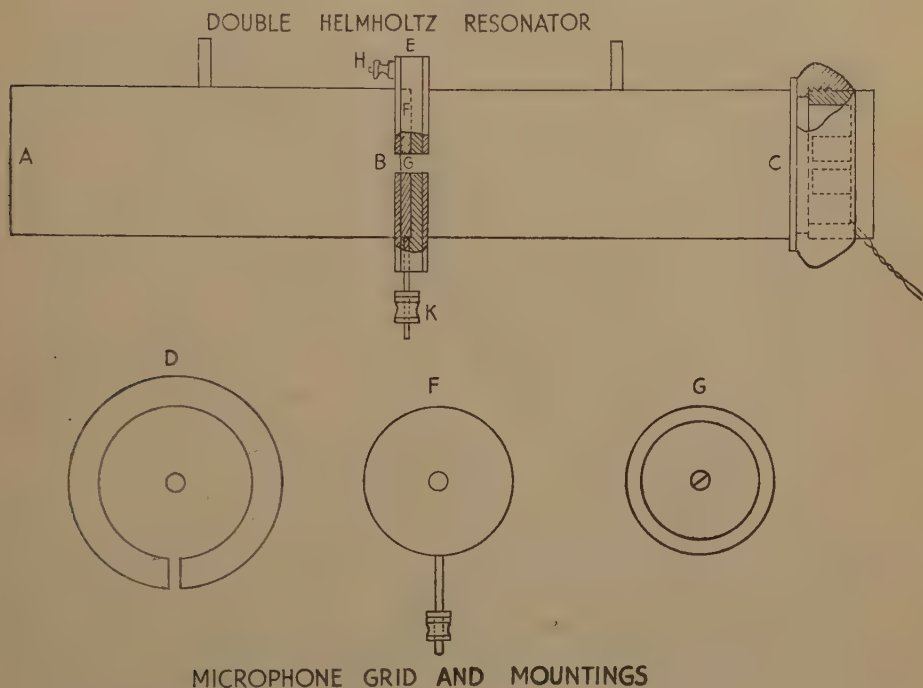
The diaphragm, behind which the usual electromagnets in their normal telephone mounting were screwed, could then serve as a source of sound actuated by the electric current of a beat-frequency oscillator.

Between the two flanges, terminating the central portions of the double resonator, an ebonite holder was inserted consisting of a disk with a shallow concentric cavity about 2 mm. deep connected to the edge of this disk by a narrow channel, as shown in the side view D and in position E. Into this cavity was fitted a flat nickel disk F slightly less than 2 mm. thick. To this was brazed a flat metal rod which exactly fitted the channel, carrying at the end a screw terminal with an insulating ebonite washer. The centre of the disk and of the the ebonite holder was drilled with a hole 6 mm. in diameter so that it could form part of a

smooth channel continuous with the circular openings in the two halves of the resonator.

The microphone of the hot-wire type consisted of a single strand of fine platinum wire, 0.0025 cm. in diameter, mounted in the window of a circular mica disk (G). The two faces of the mica were coated with thin annular silver plates on either side, forming electrodes to which the ends of the thin wire were soldered. One face of the mica plate with its electrode rested on the metal disk above described, the other presented a metal surface to the flange of one of the resonator units.

Fig. 2.



The ebonite disk, with its microphone in place, was then bolted to the two flanges of the resonator units so that they formed a rigid air-tight chamber, the halves of which were connected by a smooth cylindrical intercommunicating passage. One of the bolts clamping the units together, and the terminal of the nickel electrode, served for conveying the electric current through the platinum wire (H, K). Both resonator units were provided with narrow brass side tubes by which the various gases and vapours could be introduced.

For the purpose of testing the properties of the double resonator a second unit was made from similar tubing, but of length 21.8 cm. This alternatively could be bolted to the telephone-equipped unit with the

intervening ebonite holder, so that the same microphone in the same channel could be employed.

The function of the microphone was twofold. When heated by a small electric current it was capable of acting as a platinum resistance thermometer. The nature of its mounting and the insulation of the cavity from the outer air and any consequent draughts, made it a very satisfactory device for giving accurately the temperature of the air in the channel just at that point at which the temperature was required to be known. It presents the slightest obstruction to oscillatory air currents and, moreover, is extremely sensitive. As used with a standard Post Office bridge and a well-charged secondary cell it gives a very steady zero, and throughout the experiments under controlled conditions it gives faithful reproductions on repetition. During the conduct of the experiment, and before the sound is injected into the resonator, a measurement of temperature can be made to a fraction of a degree.

(2) *The Source of Sound and the Microphone Response.*

The sound was generated by the use of a Marconi-Ekco beat-frequency oscillator giving an open scale between 0 and 300 cycles per second, and beyond this to frequencies up to 12 kilocycles. The lower range was adequate for the experiments to be described. By means of the cathode ray oscillograph the output of the oscillator was proved to give a good simple harmonic wave form over the ranges employed. The telephone diaphragm generating the sound gave, however, a series of harmonics which at first sight might be a serious handicap. The high frequency constituents to which the resonator might respond were, however, completely eliminated by the insensitivity of the hot wire. On the other hand, the lower harmonics were exceedingly valuable as a check in the determination of frequency, for the resonance peaks occurred at points in harmonic series very accurately, and in some cases gave maxima even better to locate than the principal maximum.

A feature of the resonance curves was the close family relationship when the gases employed were of the same character, such as dry and saturated air at varying temperatures. The remarkable variation in damping and the occasional partial or almost complete suppression of some of the subsidiary peaks, when all the experimental conditions were apparently the same, might form with advantage the subject of further research. Some suggestions for these variations are made as typical curves are exhibited. The maxima for the main peaks were, however, always easy to locate, and the perfection in the shape of the resonance curves due to the complete insulation of the enclosure from external disturbances of draught, sound and temperature change, was no doubt responsible.

(3) *Measurement of Temperature.*

Throughout the whole series of experiments, the platinum thermometer was the single strand of platinum above referred to. Its resistance at

0° C. was 2.353 Ω . Its mean temperature coefficient between 0° C. and 100° C. was 0.00328 and allowance was made for the departure from the linear relation. An estimated accuracy to 0.1° C. was assumed. A standard Post Office Box was employed. A very small heating current was employed, coupled with a sensitive galvanometer in order not to raise the temperature of the gas in the channel to any appreciable extent.

(4) *Check on frequency.*

As the oscillator was driven from the mains the output was extremely steady after a certain period was allowed to elapse. The instrument was set on the reading of 50 cycles and the accuracy was assured by its synchronization with the mains frequency. For this purpose a subsidiary circuit was provided and the zero setting was adjusted by the almost complete suppression of beats. This operation was repeatedly checked and was, of course, necessary through small variations in "mains" frequency. The frequency scale was further checked against an octave of standard tuning forks. The method employed was to sound the fork in one ear, while the other received the sound from a telephone driven by the oscillator. The presence of the first harmonic of the fundamental in the telephone made it possible to check the frequency scale over two octaves, as before, by the method of beats. By this means it was hoped to get an accuracy of 0.1 cycle/sec. as a mean of several readings, and with a steady current output as shown by the output voltmeter supplied with the instrument.

(5) *Resonance Frequencies for Dry Air at Various Temperatures.*

The resonator was filled with dry air by connecting one of the inlet tubes to a drying tube while the other was connected to a slow acting exhaust pump. It was then closed up and placed in an oven electrically heated and with a good thermostatic control. Each experiment occupied two or three hours, and no observations were taken until the thermometer in the resonator gave a steady temperature reading. An accuracy of 0.1° C. was obtainable by this means. Table III. gives the resonance points observed, and the corresponding velocities of sound were calculated from the relation

$$c_t = c_0(1 + 0.00183t),$$

where c_0 is taken as 331.8 m./sec. and t is the temperature in °C.

Four sets of resonance curves were obtained. At the highest temperature the ebonite showed a tendency to soften and it was not considered desirable to continue at higher temperatures. The curves were repeated several times, observations being taken every five cycles, and when the approximate resonance point was found, the maximum was slowly passed through, upwards and downwards, and the mean obtained, the mains frequency adjustment at 50 cycles being made between each reading. Values to 0.1 cycle were assessed and were then corrected from the frequency calibration scale.

(6) *Resonance Frequencies for other Gases.*

(i.) Carbon dioxide.

Observations were made repeatedly with pure dry carbon dioxide gas at air temperature. The resonator was repeatedly exhausted by a vacuum pump in order to remove all traces of air, and when the temperature was steady a series of observations was immediately taken in order to ensure that no diffusion of air had taken place through leakage. A much lower frequency was obtained for resonance. After three hours the frequency was found to be 0.6 cycle higher, thus giving a measure of that diffusion.

(ii.) Hydrogen.

The resonance curves for hydrogen of guaranteed purity were also obtained at air temperature. The values for both these gases are included in Table III.

The values of resonance-frequency (r) and velocity (c), which in the case of these two gases is accurately known, were correlated and the ratio is shown in column 4 of the table. They are seen to be sufficiently constant to regard the velocity of sound in the gases as being proportional to that frequency. The variation from the linear relation is shown in column 5. The mean value of the constant is 1.692.

For the resonator employed therefore $c = 1.692 r$, with an accuracy of about ± 0.3 per cent.

TABLE III.

1 Gas and temp. ($^{\circ}\text{C}.$).	2 Resonance pt. (r) cycles/sec.	3 Velocity (c) (m./sec.).	4 c/r .	5 Variation from mean (%).
Air 19	202.2	343.3	1.6973	+0.32
„ 54.2	215.4	364.6	1.6926	+0.04
„ 65.0	220.2	371.7	1.6876	-0.25
„ 104.3	233.5	395.1	1.6921	+0.01
CO ₂ 18.6	158	266.9	1.6893	-0.15
H 16	3×256.5	1299	1.6924	+0.03

Mean value of c/r 1.692.

(7) *Remarks on the Resonance Curves.*

The diagram (fig. 3) shows five of the resonance curves referred to above, and the following comments are made.

For purposes of comparison the same scale of frequency is employed. The scale of ordinates was adjusted so that the resonance curves were more nearly comparable.

Curve 1 is a representative curve for carbon dioxide and is the sharpest of the whole series, indicating relatively small damping. The degree of response was actually smaller than that for air, with the same intensity of injected sound. The lower viscosity, about 83 per cent. that of air, would account to some extent for the lower damping; and the small response might be attributed to its greater density, which is 50 per cent. above that of air and to its lower thermal conductivity which is 54 per cent. that of air, and therefore would cause a reduction in temperature variation. The harmonic response at 79 cycles/sec. is definite and confirms the resonance frequency of 158 quoted in the table.

Curves 2, 3 and 4, are characteristic curves for dry air. The degree of damping was higher than might have been expected in a cavity of good resonating properties, and the resonance peaks are noticeably not so sharp as those obtained with the single resonator with open neck used by Paris and the author*, although the dimensions of orifice and volume were comparable and gave a comparable frequency. It was found that a reduction in the intensity of the injected sound led to a sharpening of the resonance peaks with, of course, a loss in response. A tendency to a lower resonance frequency maximum was also shown, but was not sufficiently definite to justify incurring the added difficulty in observing the maximum through loss of sensitivity.

The degree of damping as shown by the relative sharpnesses of the peaks becomes progressively greater as the temperature rises. This would be expected from the increase of viscosity of the air as shown by the following data:—

<i>Temperature.</i>	<i>Viscosity.</i>
20°	1.48×10^{-6} c.g.s. units.
54°	1.94×10^{-6} „ „
104°	2.2×10^{-6} „ „

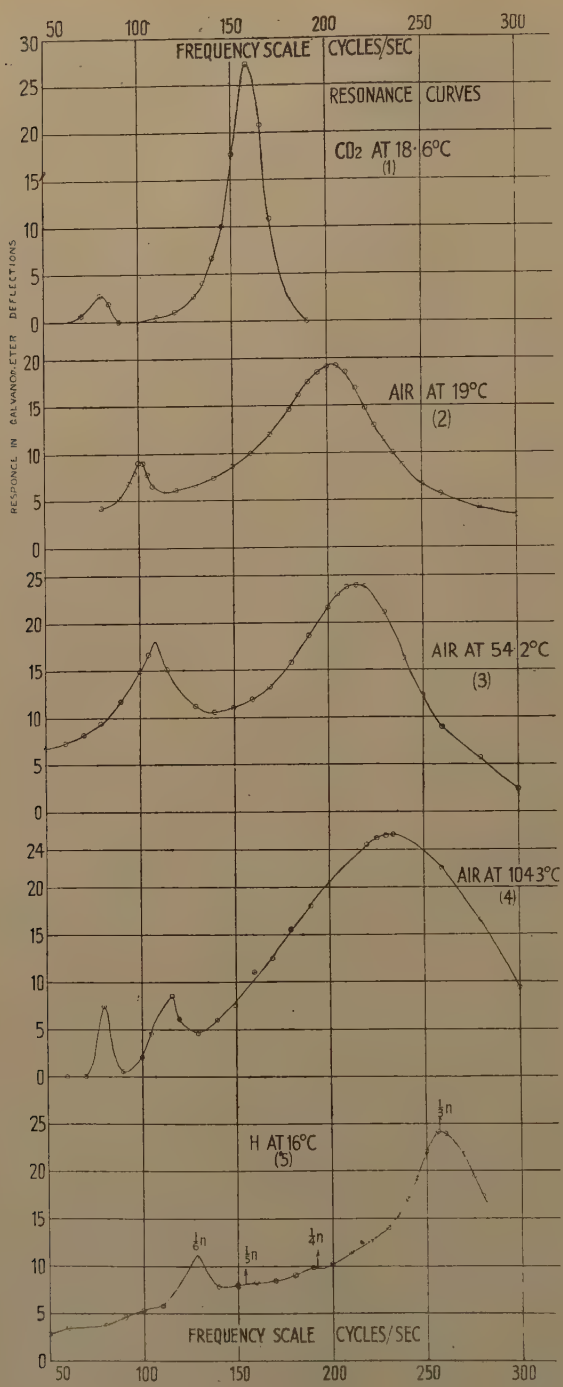
As the temperature rises the cavity evidently gradually loses its resonating properties.

The resonance peaks due to harmonics are definite for the first harmonic and are sharper than the main peak. Those due to the second harmonic of the injected frequency gradually grow, being difficult to place at air temperatures, just perceptible at 54.2°, while at 104.3° the peak almost equals that of the first harmonic. This may be due partly to the change in elastic properties of the diaphragm as the temperature rises.

Curve 5 is one of a number of similar resonance curves obtained from hydrogen. This gas was chosen because its velocity was accurately known, and was far outside the range of velocities of other gases. It

* Tucker and Paris, *loc. cit.*

Fig. 3.



could also be obtained chemically pure. The same precautions were taken in filling the resonator, which was freed from any possible leakage by the use of plasticine. The resonance points were obtained only at air temperature and no delay, subject to steady conditions of temperature, was permitted in making observations.

The response to injected sound, however, presented a remarkable change from that given by air and other gases. It was found that a response as definite as with the other gases was given when that frequency was 256.5 cycles/sec., *i. e.* one-third of that expected, but at higher frequencies no appreciable response was obtained. A slight maximum was produced at frequencies between 380 and 390 cycles, but nothing was detected in the higher frequency region which, according to the resonance frequency-velocity relation, should be at twice this frequency. Another resonance peak was obtained at approximately 130 cycles, but the intermediate frequencies in harmonic relation, of about 154 and 190 cycles/sec., were almost entirely suppressed. The evidence, therefore, as far as it goes, is that response is only obtained in any appreciable degree to sounds emitted by the diaphragm whose fundamental frequencies were one-third and one-sixth of the natural frequency of the resonator. There is no justification for assuming that the telephone diaphragm, in precisely the same mounting as with the other gases and at air temperatures, should give an output of different quality.

Reduction in sensitivity to a frequency as high as 700–800 cycles would be anticipated owing to the thermal capacity of the hot wire, but it is only due to the existence of this frequency in the impure sub-harmonics that response can be detected. This being the case, there would seem to be no justification for assuming that, for reasons of sensitivity of microphone, one sub-harmonic should be favoured more than another and that if the main resonance peak is suppressed for this reason, the others should similarly be suppressed.

The low density of hydrogen would be favourable for producing larger amplitudes—on the other hand, its high thermal conductivity, six times that of air, would reduce the temperature changes in the microphone and therefore its sensitivity. Damping due to viscosity should be less, as that for hydrogen is about half that for air at the same temperature. Whatever contributions each of these physical properties make to the behaviour of the hot wire in hydrogen, it may be stated that the curves actually obtained give a resonance peak sharper than the main peak for air, and with approximately the same amplitude.

A possible explanation for this discrimination of response may be the generation of sound by the hot wire itself, which under the influence of the injected sound suffers periodic heating and cooling, thereby acting as a thermophone*, the resultant effect being to destroy the oscillatory air currents in the neck of the resonator. A wire heated by a direct

* It is obvious that a serious investigation of the hot wire as a generator of sound is justified.

current can, when oscillatory current is imposed on it, give rise to periodic waves synchronous with the electric current and its octave. Such a secondary effect may be more distinguishable with hydrogen than with other gases owing to the greater mobility of the hydrogen molecules. This possibility suggests lines of further research, and no attempt to find a solution is made in the present paper. It is, however, fair to assume that if the hot wire were to behave with hydrogen as with other gases it would give a definite response to a sound of three times the frequency of that recorded. Using such a figure it will be seen that hydrogen falls into line with the other gases in which the resonance-frequency is directly proportional to the velocity.

§ 3. *Velocity of Sound in Vapours.*

(1) A survey of the researches on the velocity of sound in different gases shows that little work has been done on the velocity in mixtures of gases, and more particularly on air saturated with various vapours. E. Griffiths *, using supersonic sounds, has worked with mixtures of gases and indicated a method of finding the proportions of two gases from their measured velocities and the velocity in the mixture. A number of organic vapours were employed by Dixon and Greenwood † in a direct method of determining their velocity by using a long coiled tube and measuring the time of travel of a pulse throughout its length. Variations of velocity with temperature were obtained for ether and the more volatile organic liquids. In very early experiments Masson ‡ obtained the velocity of sound in water vapour, the only record of its measurement by a direct experiment.

The apparatus above described is readily adapted to the measurement of velocity of sound in air saturated with different vapours.

In the first instance it was desired to obtain the velocity in air saturated with water vapour at various temperatures. A small quantity of water was introduced into the resonator and in the course of two or three hours it was fully saturated. A slight leak was permitted to equalize the internal pressure with that of the atmosphere. Knowing the temperature, the relative pressures of the water vapour and air, the proportion of air to vapour could be estimated.

Assuming the frequency at resonance to be proportional to the velocity, the velocity of sound in the mixture was found and the value of γ , the ratio of specific heats, derived.

Dealing with a mixture of gases and assuming as an approximation their approach to the perfect gas condition, Professor G. W. Todd has derived the following relations :—

If ρ_1 and ρ_2 are the masses of the constituents per c.c.,
 p_1 and p_2 their partial pressures,

* E. Griffiths, *Proc. Phys. Soc.* xxxix. p. 300, Aug. 1927.

† Dixon and Greenwood, *Proc. Roy. Soc. A*, cv. p. 199 (1924).

‡ Masson, *Ann. Chem. and Phys.* iii. 53, p. 257 (1858).

K_1 and K_2 their specific heats at constant volume,
 γ_1 and γ_2 the ratios of their specific heats,
 n_1 and n_2 the number of degrees of freedom,
 T_{tr} and T_{to} the translational and total energies,
 γ the ratio of specific heats of the mixture,

$$\begin{aligned}\frac{T_{tr}}{T_{to}} &= \frac{3(p_1+p_2)/2\alpha}{\rho_1 K_1 + \rho_2 K_2} = \frac{3}{2}(\gamma-1), \\ \frac{3\alpha p_1}{2\rho_1 K_1} &= \frac{3}{2}(\gamma_1-1), \quad \frac{3\alpha p_2}{2\rho_2 K_2} = \frac{3}{2}(\gamma_2-1), \\ \gamma &= \frac{\rho_1 K_1 \gamma_1 + \rho_2 K_2 \gamma_2}{\rho_1 K_1 + \rho_2 K_2} = \frac{n_1 p_1 \gamma_1 + n_2 p_2 \gamma_2}{n_1 p_1 + n_2 p_2} \\ &= \frac{p_1 \gamma_1 (\gamma_2^{-1}) + p_2 \gamma_2 (\gamma_1^{-1})}{n_1 p_2 + n_2 p_1}. \quad . \quad . \quad . \quad (1)\end{aligned}$$

Also γ (for the mixture of gases) $= \frac{\rho_1 + \rho_2}{p_1 + p_2} c^2. \quad . \quad . \quad . \quad (2)$

In the case of water vapour where the vapour pressure, within the temperature ranges employed, did not exceed 27.14 cm. of mercury the departure from the perfect gas condition would probably not be greater than the experimental errors.

In order to be certain that the air was fully saturated and conditions stable, the platinum thermometer was required to give a steady zero without injecting sound, and also after this was assured, a constant deflexion should be obtained in the galvanometer when a constant sound was maintained. As a variation of deflexion near the resonance point is very definite with slight change of frequency, the twofold check of temperature and humidity assured the necessary conditions.

The resonator was placed in the oven described above, the thermostatic control being set for a given temperature. Three hours were allowed for this experiment. Owing to the nature of the thermometer mounting and its protection by the brass walls of the enclosure slight variations of oven temperature, through slight lag in the control, were not communicated to the thermometer.

The following observations of the resonance point for the mixture and the calculated corresponding velocity are shown in Table IV.

Owing to the action of heat and water vapour on the ebonite fittings which tended to soften, no attempt was made to work at higher temperatures. The velocity was calculated from the relation $c=1.692 r$.

The densities of the mixtures were determined from the known partial pressures at the temperatures quoted. The atmospheric pressure fluctuated slightly about a mean of 76 cm. of mercury during the series, and such variations produced only negligible variations of density from the mean value.

The value of γ for the mixture was derived from equation (2), and taking γ_1 , for air as 1.401, that for water vapour γ_2 was derived from

equation (1), assuming the mixture to behave as a perfect gas. How far this assumption may be regarded as reasonable can be judged by the results, for which no greater accuracy is claimed than that given by the experimental determination of the velocity of sound.

It must be noted that an error of 0.5 per cent. in the velocity of sound gives an error of 1 per cent. in the value of γ , which in the value of $\gamma-1$ may become as much as 4 per cent.

TABLE IV.

Temperature (°C.).	Resonance pt. (r) (cycles/sec.).	Velocity (c) (m./sec.).
19	203.2	343.8
38	211.6	358.1
51.6	216.9	367.0
56	220.9	373.8
63	224.8	380.4
73.5	232.0	392.6

TABLE V.

Temp. (°C.).	Vapour press. (cm. of Hg).	Density of saturated vapour ($g/c.c.$).	Velocity (c) (m./sec.).	Values of γ .	
				Mixture.	Water vapour.
19	1.65	$119.9+10^{-5}$	343.8	1.398	1.309
38	4.95	$110.6+10^{-5}$	358.1	1.396	1.336
51.6	10.01	103.2×10^{-5}	367.0	1.371	1.254
56	12.38	100.7×10^{-5}	373.8	1.386	1.329
63	17.14	96.0×10^{-5}	380.4	1.370	1.295
73.5	27.14	88.3×10^{-5}	392.6	1.344	1.274

The mean value for γ for water vapour is 1.301.

(2) may be written in the form

$$\frac{p_1+p_2}{\gamma-1} = \frac{p_1}{\gamma_1-1} + \frac{p_2}{\gamma_2-1}$$

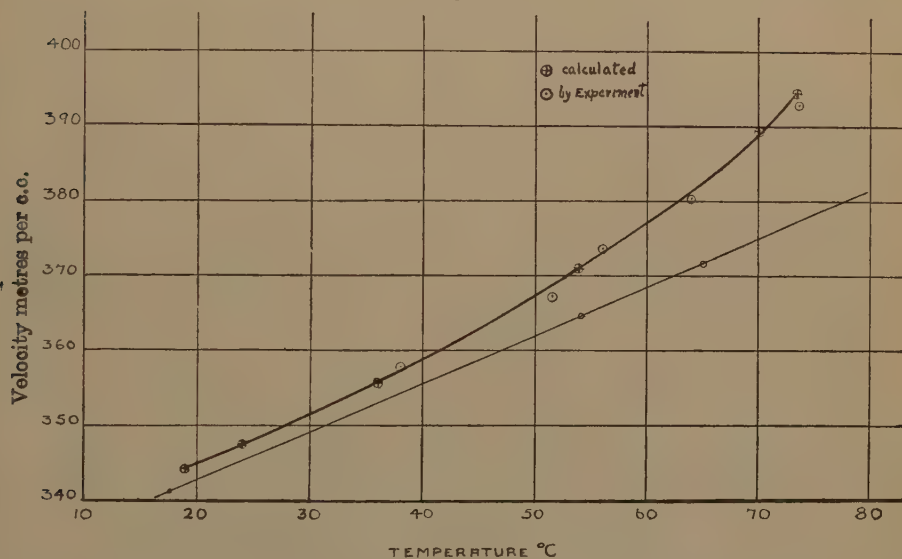
and the importance of a correct value in the velocity of sound is appreciated.

Table V. summarizes the results obtained at six different temperatures.

TABLE VI.

Temp. (°C.).	Partial press. (cm. of Hg.)		Densities (g/c.c. $\times 10^{-5}$).			Values of γ .	Velocity of sound (m./s.).
	Water vapour.	Air.	Water vapour.	Air.	Mixture.		
19	1.65	74.35	1.64	118.3	119.9	1.401	344.1
24	2.24	73.36	2.20	115.4	117.6	1.400	347.3
36	4.46	71.54	4.20	107.5	111.7	1.395	355.8
54	11.25	64.75	10.0	91.9	101.9	1.384	371.0
70	23.37	52.73	19.8	71.4	91.2	1.365	389.5
73.5	27.14	48.86	22.8	65.5	88.3	1.357	394.6

Fig. 4.



Adopting a mean value of 1.300 for water vapour, the following values for γ for the mixture of air and water vapour and the values of velocity of sound for the mixture were computed. These are shown in Table VI.

A curve was plotted connecting velocity and temperature from the values in Table VI., and the values obtained by experiment in Table V. are placed on the same diagram (fig. 4). Table VII. shows how nearly the experimental values agree with those calculated. It will be seen that the relatively large variations in γ for water vapour are not so evident in the velocity readings.

For the purpose of comparison the straight line relation connecting velocity and temperature for dry air is included in the same diagram. With a certain degree of accuracy the humidity of the atmosphere at any temperature can be derived from observed values of velocity, assuming that a linear interpolation can be employed. For this purpose, however, observations should be made with higher precision and with apparatus of improved design.

A check on the above results can be made by use of existing data.

TABLE VII.

Temperature.	19°	38°	51.6°	56°	63°	73.5°
Computed velocity.	344.1	357.7	368.6	373	380.7	394.6
Observed velocity.	343.8	358.1	367.0	373.8	380.4	392.6

Callendar's * values for γ saturated steam over the range here employed were

20°	1.287
40°	1.288
60°	1.292
80°	1.296

The mean of these gives 1.291 as against the mean value quoted of 1.301.

His value of 1.286 at 0° gives the velocity of sound in saturated steam at 0° as 402.2 m./sec. Taking the experimental value for 19° above and correcting for temperature on the basis of a coefficient of expansion of water vapour 0.00419, the value for velocity at 0° is 402.7 m./sec. A direct determination made by Masson gives a value of 401 m./sec.

§ 4. *Examination of other Saturated Vapours.*

(1) *Ether.*

The mixture of air and ether vapour provides a bigger variation of velocity of sound with temperature owing to its higher vapour pressure.

A number of experiments was conducted at air temperatures to obtain the velocity of sound in air saturated with the vapour. As before, a small quantity of the liquid, adequate to produce saturation, was introduced

* Partington and Shilling, 'Specific Heat of Gases,' p. 174.

into the resonator. The temperature of the mixture was accurately found, and the condition of saturation was found by the steadiness of the galvanometer deflexion when the output of the oscillator was kept constant. As the frequency of resonance gradually falls with growing saturation the process can be followed and the condition of complete saturation assured. A number of resonance curves was taken, the mean of several frequencies giving the maximum obtained.

The observed temperature was 17.2°C . As before, a slight leak assured equalization of the internal pressure with that of the atmosphere which was 75.3 cm. of mercury. The partial pressure of ether derived from tables was 39.5 cm. at this temperature, and the density of the mixture of air and ether vapour found to be 0.002194 gm./c.c. at normal pressure.

The mean resonance frequency was 138 c./sec., corrected for error of frequency scale to 137.1 c./sec. The velocity of sound found from the relation $c=1.692 r$ was 232.0 m./sec. From the relation $\gamma=\frac{c^2\rho}{p}$ the value for γ of the mixture was 1.1644, and substituting this in the relation

$$\frac{p}{\gamma-1} = \frac{p_1}{\gamma_1-1} + \frac{p_2}{\gamma_2-1},$$
 γ_2 for ether vapour is 1.0714, from which the velocity of sound at 17.2°C . in ether vapour is 186.6 m./sec.

The coefficient of expansion of ether vapour is not known. Assuming it to be in the region of 0.0038 the velocity at 0° would be 180.7 m./sec. The only record of velocity for pure ether vapour derived by experiment is that of Masson, who obtained the value of 179.2 m./sec. at 0°C . Leduc from thermodynamical considerations obtained a value of γ for ether of 1.08, and Dixon and Greenwood obtained 1.086 as against 1.0714 in the present paper. Using their value for γ the velocity at 0° would be 182.3 m./sec.

(2) *Acetone.*

A series of observations were made for air impregnated with acetone vapour at air temperature and the following data obtained :—

Temperature 16.5°C .

Atmospheric pressure 75.5 cm.

Mean observed resonance frequency 178.1 c./sec.

Corrected from frequency calibration curves 177.8.

Velocity of sound $1.692 \times 177.8 = 300.8$ m./sec.

The maximum vapour pressure at 16.5°C . from tables, 15.6 cm. of mercury.

Density of mixture at N.T.P., 0.001455 gm./c.c.

Value of γ for mixture 1.299.

This gives a value for γ of the pure acetone vapour 1.0386 and a velocity of sound at 0°C . of 201.2 m./sec.

No records of results from other investigators have been discovered.

(3) *Comments on the resonance curves of mixtures of air and saturated vapours.*

The same types of resonance curves were produced as those obtained for air and gases. As with dry air, at increasing temperatures the damping of resonance was progressive and was heavier than that of air freed from moisture. A series of curves was obtained when the output of the source of sound, as actuated by the acoustic oscillator, was kept constant, and some rough estimate of the increase in damping might be attempted. The response of the resonator to sub-harmonics of the true resonance frequency was very similar to that for pure air with the same prominent second peak in the case of air saturated with water vapour.

Reference has already been made to the reduction in damping when weaker sounds were employed. In spite of loss in sensitivity the maxima were distinguishable and served as a useful check. No attempt is made to explain this reduction in damping, but it is possible that the intense vibrations set up within the enclosure by loud sounds give rise to some degree of turbulence hostile to resonance.

The sub-harmonic responses for ether, acetone and carbon tetrachloride, when impregnating the air, appear to be very erratic, although in each individual case they faithfully repeat each other.

Ether gives a response to the second sub-harmonic sharp and about 75 per cent. of the main harmonic, but the first sub-harmonic response is nearly suppressed. With carbon tetrachloride, even when the air has not reached saturation, the suppression is more obvious. Acetone gives very prominent responses to first, second, and third sub-harmonics, all responses being more than 50 per cent. of the main one. The peaks, however, where they exist give resonance points in harmonic relation to the main resonance point within the degree of accuracy claimed. Whatever the cause of these phenomena, whether due to the varying quality of output of the telephone diaphragm, in all outward respects in the same condition of mounting and stimulation, the additional maxima give a valuable check on the value derived from the main maximum. It is proposed to carry out these investigations further.

(4) *Conclusion.*

It is believed that, in so far as the measurement of velocity of sound is derived from one quantity only, viz., that of frequency at the resonance condition, the method here described presents advantages over those derived from measurements of nodal distribution in pipes in addition to estimations of frequency. Such measurements, as usually carried out, also involve the use of high frequency sources of sound.

As compared with the use of pipes or resonance tubes, the use of the Helmholtz resonator, within the limiting condition of a wave-length large compared with its dimensions, has the advantage of single frequency response, and no troublesome corrections have to be applied in order to co-ordinate the results with those to be obtained from free gases. A

sufficient number of gases have been exhaustively examined to provide reliable values of velocity of sound from which an accurate calibration of the resonator can be made. Even if this were not so, the relation between velocity of sound and temperature is well enough established to give the necessary data for calibration, if the velocity of sound in one such gas is accurately known.

Control of temperature, and insulation from noise or outside disturbance render observations rapid and easy to record. Small quantities of gas or vapour can be employed, and the apparatus can readily be worked in relatively small thermostatic enclosures.

The assumptions as to the close approach to perfect gas conditions with saturated vapours is too approximate for complete treatment, but the results obtained for γ , and velocities of sound based on a knowledge of its value, accord reasonably well with data at present available. Higher accuracy could be obtained with apparatus of greater precision which would justify a more exacting research.

In conclusion, the author wishes to express his indebtedness to Professor G. W. Todd, of King's College, Newcastle, for his interest and the facilities provided, and to Professor G. R. Clemo for the supply of chemically pure gases and liquids.

XXVII. *Rational Electrodynamics*.—IV. The “Radius” of a Point Charge.

By E. A. MILNE, F.R.S.*

[Received October 20, 1942.]

79. The object of Part IV. of this series of papers is to apply the general theory of electrodynamics developed in Part III. to the interaction of two point-charges. This is the general two-body problem of electrodynamics. But, as in gravitational theory, progress is made by first considering the simpler “one-body” problem, that in which one of the bodies has such a large mass compared with the other's that it may be considered as fixed. Hence in this paper we consider two point-singularities P_1 and P_2 , of charges e_1 and e_2 , and masses m_1 and m_2 , of which m_2 is large compared with m_1 . We require the motion of e_1 about e_2 .

Our general electrodynamical theory has been constructed to be completely relativistic, in the sense that all equations are unaltered in

* Communicated by the Author.

form on changing from any one fundamental observer to any other. We can now without loss of generality simplify our problem by choosing a special observer O. Since P_2 is sufficiently massive to be considered as uninfluenced by the presence of P_1 , we may take P_2 to coincide with a fundamental particle, and we choose our observer O at this fundamental particle. Thus P_2 is at O, and $\mathbf{P}_2=0$. Then also $\mathbf{V}_2=0$, $Y_2=1$. Omitting the suffix 1 since we are considering the motion of P_1 only, we now have for the field of e_2 at e_1 , by (70) and (70'),

$$\mathbf{H} = \frac{e_2}{c} \frac{1}{Y^{\frac{1}{2}}} \left(\frac{1}{2} \mathbf{V} \wedge \frac{\mathbf{P}}{|\mathbf{P}|^3} \right), \quad \dots \quad (94)$$

$$\mathbf{E} = \frac{e_2}{Y^{\frac{1}{2}}} \frac{\mathbf{P}}{|\mathbf{P}|^3}. \quad \dots \quad (94')$$

The first of these puts in evidence the factor $\frac{1}{2}$ which our analysis has required as giving expression to the magnetic field at P_1 due in this case to the motion of P_1 , relative to the stationary charge P_2 . Both correspond to pure point-singularities "acting" according to the inverse square of the distance. Both contain a velocity factor $Y^{-\frac{1}{2}}$ or $(1-V^2/c^2)^{-\frac{1}{2}}$ which is of fundamental importance in the sequel.

80. The mechanical force on the charge e_1 is now given by formulæ (51), (51') of the preceding paper, and the resulting equations of motion are given by (52) and (52'). Using (94) and (94'), we have for the desired accurate equations of motion of P_1 ,

$$\frac{m_1^{\xi^{\frac{1}{2}}}}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(\frac{\mathbf{V}}{Y^{\frac{1}{2}}} \right) = - \frac{m_1^{\xi^{\frac{1}{2}}}}{X} \left(\mathbf{P} - \mathbf{V} \frac{Z}{Y} \right) + \frac{e_1 e_2}{t_0} \frac{Z}{Y^{3/2}} \frac{\mathbf{P} + \frac{1}{2} \mathbf{V} \wedge (\mathbf{V} \wedge \mathbf{P})/c^2}{|\mathbf{P}|^3}, \quad (95)$$

$$\frac{m_1^{\xi^{\frac{1}{2}}}}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(\frac{c}{Y^{\frac{1}{2}}} \right) = - \frac{m_1^{\xi^{\frac{1}{2}}}}{X} \left(ct - c \frac{Z}{Y} \right) + \frac{e_1 e_2}{t_0} \frac{Z}{Y^{3/2}} \frac{\mathbf{P} \cdot \mathbf{V}}{c |\mathbf{P}|^3}. \quad \dots \quad (95')$$

Of these equations, (95') is the scalar fourth equation corresponding to the three equations comprised in the single vector equation (95): (95') follows from (95) on scalar multiplication by \mathbf{V} .

81. Several comments on these equations may be made. In the first place they should be *exact*. No term has been omitted which might correspond to a supposed "reaction of emitted radiation" on the moving charge e_1 . In this connexion we note the absence of any mention of the epoch t_2 at the charge e_2 ; the t in (95) and (95') is the epoch at the charge e_1 . There is thus no question of our having to relate t_2 and t_1 , and so there can be no radiative interaction between e_2 and e_1 .

In the second place, this exact formulation holds good in t -measure, in the private Euclidean space of the observer O at e_2 . In t -measure, the first term on the right-hand side of (95) represents the natural motion that e_1 would have if free; it is transformed away when we pass to τ -measure. In this t -measure, $m_1^{\xi^{\frac{1}{2}}}$ represents the mass of the particle as in motion. It will be noted that it appears *outside* the operator d/dt . This is a consequence of the form of equations (51), (51') defining

the external force, which contained a term taking proper account of the change of mass with velocity. The term in $d(m\xi^{\frac{1}{2}})/dt$ has disappeared due to a cancellation from the two sides of (95)—it is not that we have omitted to take into account the change of mass with velocity under the sign d/dt . The particular way in which the mass appears in (95) is a consequence of our having attended carefully to the dynamics of “external force” due to a charge, and (95) answers the old question as to whether the “external force” includes the effect of the rate of change of mass with changing velocity. Equations (95), (95') are end-products of a complex sequence of dynamical reasoning, and could not be written down “from sight.”

In the third place we note the factor $Y^{-3/2}$ in the electromagnetic terms. It arises from the product of three distinct factors $Y^{-1/2}$. One of these factors we have already noted in the formulæ for the magnetic and electric intensities (94) and (94'). A second arises from the need for the occurrence of the 4-vector $(V/Y^{\frac{1}{2}}, c/Y^{\frac{1}{2}})$ in obtaining the ponderomotive force due to the intensities by combination with the velocity. A third arises in conjunction with Z as the scalar factor $Z/Y^{\frac{1}{2}}$ which we introduced in §54 in order to arrive at the rate-of-work formula (50). These three factors ensure that our formulæ are valid for high velocities.

82. It is to be noted that it would be foreign to our method to deal with the motion by use of the supposed “laws” of conservation of energy and conservation of angular momentum. It is precisely our object to enquire whether these “laws” exist as consequences of our equations of motion, and if so, what form they take. We therefore proceed now to obtain two first integrals of (95) and (95') and compare them with the supposed “laws” of conservation.

83. In the present dynamics, kinetic energy $mc^2\xi^{\frac{1}{2}}$ is a 4-scalar, not as in Einstein's dynamics the time-component of a 4-vector. Thus (95'), though resembling an energy equation, is not actually one. As in paper II., we obtain a true energy-equation by multiplying (95) scalarly by the vector

$$\frac{\mathbf{V}}{Y^{\frac{1}{2}}} - \mathbf{P} \frac{Y^{\frac{1}{2}}}{Z},$$

multiplying (95') by the corresponding time-component

$$\frac{c}{Y^{\frac{1}{2}}} - ct \frac{Y^{\frac{1}{2}}}{Z},$$

and subtracting. On reduction we find

$$\frac{d}{dt}(mc^2\xi^{\frac{1}{2}}) = \frac{e_1 e_2}{t_0} \frac{1}{Y^{\frac{1}{2}}} \left[t(\mathbf{P} \cdot \mathbf{V}) - \mathbf{P}^2 + \frac{1}{2} \frac{(\mathbf{P} \wedge \mathbf{V})^2}{c^2} \right] \frac{1}{|\mathbf{P}|^3}, \quad \dots \quad (96)$$

a single factor Y^{-1} surviving on the right-hand side.

84. It is important to make sure that this energy-rate formula is consistent with the general theory of Part III. This gives here

$$\frac{d}{dt}[mc^2\xi^{\frac{1}{2}} + \Phi] = \dot{\mathbf{V}} \cdot \frac{\partial \Phi}{\partial \mathbf{V}}, \quad \dots \quad (97)$$

where the electrical energy Φ is given by

$$\begin{aligned}\Phi &= \frac{1}{2} \frac{e_1 e_2}{ct_0} \left[\frac{Z_1 Z_2 + Z_{12} Z_{21}}{Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}} (X_{12}^{\frac{1}{2}} - X_1 X_2)^{\frac{1}{2}}} \right]_{\mathbf{P}_1=0, \mathbf{v}_1=0} \\ &= \frac{e_1 e_2}{t_0} \frac{t - \frac{1}{2}(\mathbf{P} \cdot \mathbf{V})/c^2}{Y^{\frac{1}{2}} |\mathbf{P}|} \dots \dots \dots (98)\end{aligned}$$

Writing (97) in the form

$$\frac{d}{dt}(mc^2 \xi^{\frac{1}{2}}) = -\frac{\partial \Phi}{\partial t} - \mathbf{V} \cdot \frac{\partial \Phi}{\partial \mathbf{P}}, \dots \dots \dots (97')$$

and introducing for Φ from (98), we easily recover (96).

Equation (96) may now be written

$$Y^{\frac{1}{2}} \frac{d}{dt}(mc^2 \xi^{\frac{1}{2}}) = -e_1 e_2 \frac{d}{dt} \left[\frac{t}{t_0 |\mathbf{P}|} \right] + \frac{e_1 e_2}{t_0} \frac{\frac{1}{2}(\mathbf{P} \Delta \mathbf{V})^{\frac{1}{2}}}{c^2 |\mathbf{P}|^3} \dots \dots \dots (99)$$

So far our analysis is exact, but here for the first time we approximate. It is clear that even for $|\mathbf{V}| \sim c$, the last term on the right-hand side of (99) is utterly negligible compared with the first term, owing to $|\mathbf{P}|$ being negligible compared with ct_0 ; this is more readily seen from (96). Hence we have, to a very close approximation,

$$Y^{\frac{1}{2}} \frac{d\xi^{\frac{1}{2}}}{dt} = -\frac{e_1 e_2}{mc^2} \frac{d}{dt} \left[\frac{t}{t_0 |\mathbf{P}|} \right] \dots \dots \dots (99')$$

85. We now transform, since the problem has now become a dynamical one, to our more familiar τ -measures. We re-graduate observers' clocks from t to τ by the re-graduation formula

$$\tau = t_0 \log(t/t_0) + t_0,$$

so that

$$dt/t = d\tau/t_0.$$

Co-ordinates $|\mathbf{P}|$ and t , of an event, become λ and τ , where

$$\begin{aligned}t &= t_0 e^{\frac{\tau-t_0}{t_0}} \cosh \frac{\lambda}{ct_0}, \\ |\mathbf{P}| &= ct_0 e^{\frac{\tau-t_0}{t_0}} \sinh \frac{\lambda}{ct_0}.\end{aligned}$$

In connexion with these new co-ordinates, we change over from the private space

$$de^2 = dx^2 + dy^2 + dz^2,$$

and space-time

$$ds^2 = dt^2 - de^2/c^2,$$

to the public hyperbolic space $d\epsilon^2$, given by (32), and space-time $d\sigma^2$, given by

$$d\sigma^2 = d\tau^2 - d\epsilon^2/c^2.$$

The relation between the two is given by

$$ds = e^{\frac{\tau-t_0}{t_0}} d\sigma,$$

so that they are indistinguishable at the present epoch, but it is worth while giving the exact transformation formula for ξ^\dagger . We have

$$\frac{dX^\dagger}{dt} = \frac{Z}{X^\dagger},$$

and if v denotes the speed in the new space,

$$d\sigma = d\tau \left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}.$$

Hence

$$\xi^\dagger = \frac{Z}{X^\dagger Y^\dagger} = \frac{dX^\dagger}{Y^\dagger dt}.$$

But

$$X^\dagger = t_0 e^{\frac{r-t_0}{t_0}},$$

and

$$Y^\dagger dt = ds = e^{\frac{r-t_0}{t_0}} d\sigma.$$

Hence

$$\xi^\dagger = \frac{d\tau}{d\sigma} = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}}.$$

Moreover, locally,

$$\mathbf{v} = \frac{d\mathbf{P}}{dt} = \frac{d}{dt} \left[\frac{t}{t_0} \mathbf{\Pi} \right] = \frac{\mathbf{\Pi}}{t_0} + \frac{d\mathbf{\Pi}}{d\tau} = \frac{\mathbf{\Pi}}{t_0} + \mathbf{v},$$

so that save at cosmically important distances \mathbf{v} is indistinguishable from \mathbf{V} . Hence Y reduces to $1 - v^2/c^2$. Hence (99') becomes

$$(1 - v^2/c^2)^{\frac{1}{2}} \frac{d}{dt} \left[\frac{1}{(1 - v^2/c^2)^{\frac{1}{2}}} \right] = - \frac{e_1 e_2}{mc^2} \frac{d}{d\tau} \left[\frac{1}{|\mathbf{\Pi}|} \right]. \quad (100)$$

86. We now write r for the polar co-ordinate $|\mathbf{\Pi}|$ and integrate, obtaining

$$\log \frac{1}{(1 - v^2/c^2)^{\frac{1}{2}}} + \frac{e_1 e_2}{mc^2} \frac{1}{r} = \text{const.}$$

Putting the constant of integration equal to $\log(1 + W/mc^2)$, we have, finally,

$$\frac{1}{(1 - v^2/c^2)^{\frac{1}{2}}} = \left(1 + \frac{W}{mc^2}\right) \exp\left(-\frac{e_1 e_2}{mc^2} \frac{1}{r}\right). \quad (101)$$

87. When v is small compared with c , and r large compared with $|e_1 e_2|/mc^2$, (101) reduces approximately to

$$\frac{1}{2}mv^2 + \frac{e_1 e_2}{r} = W, \quad (102)$$

so that in these circumstances W is what we ordinarily call the energy, and the energy is constant. Accordingly (101) will be called the energy-integral.

88. But the actual energy on the present theory is

$$mc^2 \xi^\dagger + \Phi, \quad (103)$$

or in τ -measure, by (98), to a close approximation,

$$\frac{mc^2}{(1-v^2/c^2)^{\frac{1}{2}}} + \frac{e_1 e_2}{r} \cdot \frac{1}{(1-v^2/c^2)^{\frac{1}{2}}},$$

or, say, removing the rest-energy,

$$W' = \frac{1}{(1-v^2/c^2)^{\frac{1}{2}}} \left[mc^2 + \frac{e_1 e_2}{r} \right] - mc^2, \quad . \quad . \quad . \quad (104)$$

or by (101),

$$W' = mc^2 \left[\left(1 + \frac{W}{mc^2} \right) \left(1 + \frac{e_1 e_2}{mc^2 r} \right) \exp \left(- \frac{e_1 e_2}{mc^2 r} \right) - 1 \right]. \quad . \quad . \quad (105)$$

Thus W' is not exactly constant for a non-circular path as soon as r becomes comparable with $|e_1 e_2|/mc^2$. The energy W' is indeed more nearly constant according to (105) than the crude non-relativistic formula customarily taken for the energy, namely,

$$W'' = mc^2 \left[\frac{1}{(1-v^2/c^2)^{\frac{1}{2}}} - 1 \right] + \frac{e_1 e_2}{r}, \quad . \quad . \quad . \quad (106)$$

which gives on use of (101)

$$W'' = mc^2 \left[\left(1 + \frac{W}{mc^2} \right) \exp \left(- \frac{e_1 e_2}{mc^2 r} \right) + \frac{e_1 e_2}{mc^2 r} - 1 \right]. \quad . \quad . \quad (107)$$

Thus whilst W' differs from W only by a term in $(e_1 e_2/mc^2 r)^2$, W'' differs from W by a term in $(e_1 e_2/mc^2 r)$.

89. The important point following from (105) is that the energy is no longer a *constant of the motion*. The dynamical behaviour of the moving charge is characterized by a constant W , which ceases to represent the energy so soon as the distance between the charges becomes comparable with $|e_1 e_2|/mc^2$. It is not the inverse square law which fails, but the high-speed dynamics which alters. We must beware of hypo-statizing "energy," of regarding it as a thing in itself. For example, it is quite invalid to ask where the energy has gone to when it is no longer constant. We are accustomed to translating our dynamical integrals into words, and it is not easy to change our phraseology into one appropriate to a new high-speed dynamics in which the constant of the motion does not represent what we ordinarily mean by energy. Energy is a construct, and its conservation is in any given case a matter for investigation; in our example it ceases to play the part of a constant of the motion as soon as r becomes comparable with $|e_1 e_2|/mc^2$.

Now $|e_1 e_2|/mc^2$ is the classical "radius of the electron." We thus see that although our charges are pure singularities, new phenomena come in at distances comparable with the classical "radius of the electron." That electrodynamics contains e^2/mc^2 as a fundamental length has been surmised recently by several investigators, but as an *ad hoc* assumption. In the present work it has appeared, naturally and inevitably, as a consequence of our rational formulation of a relativistic electrodynamics in company with a relativistic dynamics.

90. We proceed to examine angular momentum. We have to avoid assuming it to be conserved ; instead we have to examine its behaviour.

Multiplying the equation of motion (95) vectorially by \mathbf{P} , it may be written, in t -measure

$$\frac{m\xi^{\frac{1}{2}}}{Y^{\frac{1}{2}}} \frac{d}{dt} \left(\frac{\mathbf{P} \wedge \bar{\mathbf{V}}}{Y^{\frac{1}{2}}} \right) = \frac{m\xi^{\frac{1}{2}}}{X} \frac{Z}{Y} (\mathbf{P} \wedge \mathbf{V}) + \frac{e_1 e_2}{t_0} \frac{Z}{Y^{\frac{3}{2}}} \frac{\frac{1}{2} (\mathbf{P} \wedge \mathbf{V}) (\mathbf{P} \cdot \mathbf{V})}{c^2 |\mathbf{P}|^3}, \quad (108)$$

which again may be written, on dividing out by $\xi^{\frac{1}{2}}/Y^{\frac{1}{2}}$,

$$\frac{d}{dt} \left[\frac{\mathbf{P} \wedge \mathbf{V}}{Y^{\frac{1}{2}}} \right] - \frac{\mathbf{P} \wedge \mathbf{V}}{Y^{\frac{1}{2}}} \frac{Z}{X} = - \frac{\mathbf{P} \wedge \mathbf{V}}{Y^{\frac{1}{2}}} \frac{\frac{1}{2} e_1 e_2}{mc^2} \frac{X^{\frac{1}{2}}}{t_0} \frac{d}{dt} \left(\frac{1}{|\mathbf{P}|} \right).$$

Now, since

$$\frac{dX}{dt} = 2Z,$$

the left-hand side becomes a complete differential on multiplying by $X^{-\frac{1}{2}}$. Thus

$$\frac{d}{dt} \left[\frac{\mathbf{P} \wedge \mathbf{V}}{X^{\frac{1}{2}} Y^{\frac{1}{2}}} \right] = - \frac{\mathbf{P} \wedge \mathbf{V}}{X^{\frac{1}{2}} Y^{\frac{1}{2}}} \frac{\frac{1}{2} e_1 e_2}{mc^2} \frac{X^{\frac{1}{2}}}{t_0} \frac{d}{dt} \left(\frac{1}{|\mathbf{P}|} \right). \quad (109)$$

The integral of this vector differential equation is

$$\frac{\mathbf{P} \wedge \mathbf{V}}{X^{\frac{1}{2}} Y^{\frac{1}{2}}} = \text{const.} \exp \left[- \frac{\frac{1}{2} e_1 e_2}{mc^2} \int_{t_0}^t \frac{X^{\frac{1}{2}}}{t_0} \frac{d}{dt} \left(\frac{1}{|\mathbf{P}|} \right) dt \right]. \quad (110)$$

At the present epoch t_0 , the integral is very closely

$$\frac{t}{t_0} \cdot \frac{1}{|\mathbf{P}|},$$

or in τ -measure, $1/r$. The transform of the left-hand side to τ -measure follows from

$$\frac{\mathbf{P} \wedge \mathbf{V}}{X^{\frac{1}{2}} Y^{\frac{1}{2}}} = \frac{\mathbf{P} \wedge \left(\frac{\mathbf{P}}{t} + \frac{d\mathbf{\Pi}}{d\tau} \right)}{X^{\frac{1}{2}} Y^{\frac{1}{2}}} \sim \frac{\frac{t}{t_0} \mathbf{\Pi} \wedge \frac{d\mathbf{\Pi}}{d\tau}}{t(1-v^2/c^2)^{\frac{1}{2}}} \sim \frac{\mathbf{\Pi} \wedge \mathbf{v}}{t_0(1-v^2/c^2)^{\frac{1}{2}}}.$$

We hence obtain what we shall call the angular-momentum integral in τ -measure,

$$\frac{m\mathbf{\Pi} \wedge \frac{d\mathbf{\Pi}}{d\tau}}{(1-v^2/c^2)^{\frac{1}{2}}} = \text{const.} \exp \left[- \frac{\frac{1}{2} e_1 e_2}{mc^2} \frac{1}{r} \right]. \quad (111)$$

In scalar polar co-ordinates (r, θ) , this may be written

$$\frac{mr^2 d\theta/d\tau}{(1-v^2/c^2)^{\frac{1}{2}}} = H \exp \left[- \frac{\frac{1}{2} e_1 e_2}{mc^2} \frac{1}{r} \right], \quad (112)$$

where

$$v^2 = \left(\frac{dr}{d\tau} \right)^2 + r^2 \left(\frac{d\theta}{d\tau} \right)^2,$$

and H (not to be confused with magnetic intensity \mathbf{H}) is a constant of the motion.

91. When r is large compared with $\frac{1}{2}|e_1e_2|/mc^2$, the angular momentum

$$\frac{mr^2 d\theta/d\tau}{(1-v^2/c^2)^{\frac{1}{2}}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (113)$$

remains constant and equal to H . But when r is comparable with $\frac{1}{2}|e_1e_2|/mc^2$, the classical "radius of the electron," the angular momentum varies with r . This again occurs not as a consequence of any failure of the inverse square law, but as a consequence of our properly taking into account the magnetic effect of the relative motion and the high-speed dynamics. We see that angular momentum, like energy, is not a constant of the motion, and we must again beware of hypostatizing angular momentum as something having a real existence. It is a mere construct, like energy, and we must not be surprised that its observed constancy in larger-scale phenomena no longer holds good at the high-speeds which occur at short distances.

92. It may be mentioned in passing that our rational t -dynamics shows the actual origin of angular momentum, apart from the modification due to the electromagnetic forces, as due to the general pull of the cosmos. If we put e_1 or $e_2=0$ in (110), we have

$$\frac{|\mathbf{P} \Delta \mathbf{V}|}{Y^{\frac{1}{2}}} \propto X^{\frac{1}{2}} \sim t,$$

so that angular momentum on the t -scale is approximately proportional to t . On the τ -scale, when the charges are zero, angular momentum is constant but proportional to t_0 , the present epoch reckoned from the natural zero. This accounts for the existence of spinning systems in so many natural phenomena. The physical reason is that to the moving particle, the apparent centre of attraction is in the direction of the velocity relation to its surroundings, $\mathbf{V}-\mathbf{P}/t$, and not at the observer O . Thus angular momentum grows with epoch, on the t -scale.

93. It was suggested by Bohr long ago that the integrals of energy and angular momentum must "break down" at distances comparable with the classical radius of the electron, though I am not aware that he embodied the idea in precise analysis. Our present rational electrodynamics, without any special assumptions, and starting only with point singularities acting according to a Coulomb law, has presented us with the exact integrals of the motion contained in (101) and (112), by which it appears as if a point-charge ceased to behave as a Coulomb centre of force at the distance $|e_1e_2|/mc^2$. But as we now see, this is not due to any "structure" in the point-charge, but arises purely from the dynamics of interaction.

94. The present analysis shows a distinction between the energy integral and the angular momentum integral in that the one arises solely from the *electric* intensity between the two charges, the other solely from the *magnetic* intensity. The former involves the length $|e_1e_2|/mc^2$,

But since $x=u(e_1e_2/mc^2)$, (119) may be written

$$u + \frac{d^2u}{d\theta^2} = -\frac{1}{2} \frac{me_1e_2}{H^2} \left[\left(1 + \frac{W}{mc^2} \right)^2 e^{-x} + e^x \right]. \quad (121)$$

The actual differential equation of the orbit (121) will be identical with the non-relativistic equation (120) if we choose

$$f = \frac{1}{2} \frac{e_1e_2}{r^2} \left[\left(1 + \frac{W}{mc^2} \right)^2 e^{-\frac{e_1e_2}{mc^2} \frac{1}{r}} + e^{\frac{e_1e_2}{mc^2} \frac{1}{r}} \right]. \quad (122)$$

This, then, is the *apparent force* between the two point-charges on Newtonian mechanics. We must repeat: we have assumed only a Coulomb force between the point-charges; it is the dynamics which has given rise to the apparent force f .

97. We have the following cases of (122). When $|W| \ll mc^2$ and $r \gg |e_1e_2|/mc^2$, we have

$$f \sim \frac{e_1e_2}{r^2}, \quad (123)$$

the ordinary Coulomb law. When r is comparable with $|e_1e_2|/mc^2$, we need the two-term formula (122). When r is small compared with $|e_1e_2|/mc^2$, we have two cases to consider, according as e_1 and e_2 are of like signs or opposite signs. If e_1, e_2 are of *like* signs, and $r \ll |e_1e_2|/mc^2$, (122) becomes the *repulsion*

$$f \sim \frac{1}{2} \frac{e_1e_2}{r^2} e^{\frac{e_1e_2}{mc^2} \frac{1}{r}}. \quad (124)$$

If e_1, e_2 are of *opposite* signs, (122) becomes the *attraction*

$$f \sim \frac{1}{2} \frac{e_1e_2}{r^2} \left(1 + \frac{W}{mc^2} \right)^2 e^{-\frac{e_1e_2}{mc^2} \frac{1}{r}}. \quad (125)$$

It is clear that the apparent force of interaction, which follows an inverse square law at distances large compared with $|e_1e_2|/mc^2$, suffers a tremendous increase as r passes through values comparable with $|e_1e_2|/mc^2$. Further, it appears that like charges and unlike charges behave differently.

98. The meaning in Newtonian mechanics of the constant W in (122) can be found as follows. Insert for f from (122) in (120) and integrate from $u=0$ ($r=\infty$) to u . Let U be the Newtonian energy-constant of the orbit. This is what we ordinarily call the "binding energy." Then we get

$$\begin{aligned} \frac{1}{2} \frac{H^2}{m} \left[u^2 + \left(\frac{du}{d\theta} \right)^2 \right] &= U - \int_0^u \frac{f}{u^2} du \\ &= U + \frac{1}{2} mc^2 \left[\left(1 + \frac{W}{mc^2} \right)^2 e^{-x} - e^x \right] - \frac{1}{2} mc^2 \left[\left(1 + \frac{W}{mc^2} \right)^2 - 1 \right]. \end{aligned}$$

Comparing with (118), since $x = u(e_1 e_2 / mc^2)$, we get

$$U = \frac{1}{2} mc^2 \left[\left(1 + \frac{W}{mc^2} \right)^2 - 1 \right],$$

or
$$U = W \left(1 + \frac{1}{2} \frac{W}{mc^2} \right) \dots \dots \dots (126)$$

The constant W may be positive or negative; the only condition it is subject to is $W > -mc^2$, by (101). Thus in the case of unlike charges, the attraction may be written at "short" distances, from (125),

$$f \sim \frac{1}{2} \frac{e_1 e_2}{r^2} \left(1 + \frac{U}{\frac{1}{2} mc^2} \right) e^{-\frac{e_1 e_2}{mc^2} \frac{1}{r}}, \dots \dots \dots (127)$$

where U is the binding energy calculated non-relativistically.

99. The forms of these various formulæ depend intimately on the coefficient $\frac{1}{2}$ in the magnetic term for the interaction. Were it not present we should not arrive at the remarkable formula (122) for f , which brings in apparent non-Coulomb forces at distances comparable with $|e_1 e_2| / mc^2$ whatever the sign of $e_1 e_2$.

100. Formulæ of the type of (122) have often been suggested empirically, but in the present work (122) has emerged without any specific assumptions whatever. It is a consequence of the investigation of the possible nature of forces associated with singularities.

101. It is tempting to go beyond our present deductive theory of point-charges and apply the result to actual electrons, of mass m_e and charge $-e$, and protons, of mass m_H and charge $+e$. It is obvious that our results should hold good for an electron in the vicinity of a massive nucleus, since the nucleus may be supposed fixed; for the reaction between an electron and another electron, we should need some kind of reduced mass instead of m_e ; for a proton near a massive positive nucleus we should find the small radius $Ze^2 / m_H c^2$. It is clear that the attraction of an electron towards a positive massive nucleus is enormously increased over the inverse square attraction at very close distances, and so a cluster of electrons in a nucleus will have strong cohering effect.

102. These results on the apparent radius of a point-charge and the apparent modification of the Coulomb field between point-charges, owing to the form of the high-speed dynamics, are so promising that in the next section of this series of papers we shall apply them to discuss the energy levels of the orbits of a negatively charged electron in the vicinity of a positively charged proton, following Bohr.

XXVIII. *Rational Electrodynamics.—V. The Neutron and Nuclear Dynamics.*

By E. A. MILNE, F.R.S.*

[Received October 20, 1942.]

103. In this paper I propose to examine further, in the light of the present rational electrodynamics, the interaction of two equal point-changes of opposite signs, of which one is much more massive than the other. *Classical* theory gives for this interaction a set of Keplerian orbits which, when quantised, yield the Bohr theory of the hydrogen atom. It has long been a lacuna that there is no corresponding classical theory of the neutron, *i. e.*, a theory which, classical in form, yields a system of zero net charge and of very small radius. We shall find that our present theory of electrodynamics yields such a theory without special assumptions, on precisely the same footing as the Bohr orbits of the hydrogen atom.

104. In the theory of the preceding paper, we now take $e_2 = +\epsilon$, $e_1 = -\epsilon$, where ϵ is the positive number measuring the electronic charge. We take m to denote the electronic mass, and we identify the positive charge with a proton. This is sufficiently massive compared with an electron for us to be able to apply the theory of the preceding paper, in which e_2 was taken as permanently at rest at the observer O.

We now change the sign of x , and write

$$x = + \frac{\epsilon^2}{mc^2} \frac{1}{r}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (128)$$

The integrals of energy and angular momentum obtained previously may now be written

$$\frac{1}{(1-v^2/c^2)^{\frac{1}{2}}} = \left(1 + \frac{W}{mc^2}\right) e^x, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (129)$$

$$\frac{mr^2 d\theta/d\tau}{(1-v^2/c^2)^{\frac{1}{2}}} = He^{\frac{1}{2}x}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (130)$$

whilst the differential equation of the orbit, obtained by eliminating $d\tau$ between these two integrals, is now

$$\left[\left(1 + \frac{W}{mc^2}\right)^2 e^x - e^{-x} \right] \frac{\epsilon^4}{H^2 c^2} = x^2 + \left(\frac{dx}{d\theta}\right)^2. \quad . \quad . \quad . \quad . \quad (131)$$

* Communicated by the Author,

105. When $r \gg \epsilon^2/mc^2$, so that x is small, and $v \ll c$, these equations yield as a first approximation Keplerian orbits. The Bohr theory of the hydrogen atom follows on putting

$$H = n\hbar/2\pi. \quad . \quad . \quad . \quad . \quad . \quad . \quad (132)$$

We wish our present investigation to run as parallel as possible to this well-known theory, and though it is no part of our present aim to develop quantum theory, we shall replace H in (131) by expression (132). In adopting (132), where n is integral, we are departing from our present strictly deductive scheme, and making a special assumption, a procedure entirely foreign to our proper purpose. But we wish to see what modifications the integrals (129) and (130) imply in the standard Bohr theory, and so we use a quantum value for H . No essential part of our analysis is modified by this restriction: it merely wears a more familiar air.

106. Equation (131), which is a first integral of the orbit, may now be written in the form

$$\left(1 + \frac{W}{mc^2}\right)^2 e^x - e^{-x} = \frac{n^2}{\alpha^2} \left[x^2 + \left(\frac{dx}{d\theta}\right)^2 \right], \quad . \quad . \quad . \quad . \quad (133)$$

where
$$\alpha = \frac{2\pi\epsilon^2}{\hbar c} \sim \frac{1}{137}. \quad . \quad . \quad . \quad . \quad . \quad (134)$$

α is the fine-structure constant. It is very satisfactory that this constant should appear explicitly in the equation of the orbits.

107. The apses of the orbits are given by putting $dx/d\theta = 0$, and are accordingly the roots of the equation

$$\left(1 + \frac{W}{mc^2}\right)^2 e^x - e^{-x} = \frac{n^2}{\alpha^2} x^2. \quad . \quad . \quad . \quad . \quad . \quad (135)$$

108. Now consider the *circular orbits*. For such orbits, x is constant, and therefore two roots of (135) must coincide. Differentiating with regard to x , we see that the radius of a circular orbit is also a root of

$$\left(1 + \frac{W}{mc^2}\right)^2 e^x + e^{-x} = 2 \frac{n^2}{\alpha^2} x. \quad . \quad . \quad . \quad . \quad . \quad (136)$$

Equations (135) and (136) form a simultaneous pair for the two unknowns x and W . Eliminating W , we see that the radius of a circular orbit is a root of

$$e^{-x} = \frac{n^2}{\alpha^2} \left(x - \frac{1}{2}x^2\right). \quad . \quad . \quad . \quad . \quad . \quad (137)$$

The corresponding value of W is then given by

$$1 + \frac{W}{mc^2} = \frac{n^2}{\alpha^2} x \left(1 - \frac{1}{4}x^2\right)^{\frac{1}{2}} = e^{-x} \left(\frac{1 + \frac{1}{2}x}{1 - \frac{1}{2}x}\right)^{\frac{1}{2}}, \quad . \quad . \quad . \quad . \quad (138)$$

and the energy itself is given, by (105), by

$$\begin{aligned} W' &= mc^2 \left[\left(1 + \frac{W}{mc^2} \right) (1-x)e^x - 1 \right], \\ &= mc^2 \left[(1-x) \left(\frac{1+\frac{1}{2}x}{1-\frac{1}{2}x} \right)^{\frac{1}{2}} - 1 \right]. \end{aligned} \quad (139)$$

109. We consider the roots of (137). The order of magnitude of these roots is governed by the circumstances that the fine-structure constant α is small compared with unity. The remarkable result then follows that (137) has *two and only two roots*, one in the neighbourhood of $x=0$, the other in the neighbourhood of $x=2$. There are, therefore, two sets of circular orbits. We shall find that the root near $x=0$ corresponds to the circular Bohr orbits of the hydrogen atom. The root near $x=2$ foreshadows the existence of a system of much smaller radius, which we shall identify with the neutron. We consider these roots in turn.

110. *Root near $x=0$.* The root near $x=0$ is at once found to have for its expansion

$$x = \frac{\alpha^2}{n^2} \left[1 - \frac{1}{2} \frac{\alpha^2}{n^2} + \frac{1}{2} \frac{\alpha^4}{n^4} + \dots \right]. \quad (140)$$

The leading term gives for the corresponding radius

$$r = \frac{\epsilon^2}{mc^2} \frac{1}{x} \sim \frac{\epsilon^2}{mc^2} \frac{n^2}{\alpha^2} = \frac{n^2 \hbar^2}{4\pi^2 m \epsilon^2}. \quad (141)$$

When n is integral, these are the Bohr circular orbits of the hydrogen atom. Expanding (138), the corresponding value of the constant W is

$$\begin{aligned} W &= mc^2 \left(-\frac{1}{2}x + \frac{1}{8}x^2 + \dots \right), \\ &= -\frac{1}{2} mc^2 \frac{\alpha^2}{n^2} \left(1 - \frac{3}{4} \frac{\alpha^2}{n^2} + \dots \right), \end{aligned} \quad (142)$$

and by (139), the corresponding value of the energy W' is

$$\begin{aligned} W' &= mc^2 \left(-\frac{1}{2}x - \frac{3}{8}x^2 - \dots \right) \\ &= -\frac{1}{2} mc^2 \frac{\alpha^2}{n^2} \left(1 + \frac{1}{4} \frac{\alpha^2}{n^2} + \dots \right). \end{aligned} \quad (143)$$

The leading term thus gives

$$W = -\frac{1}{2} mc^2 \frac{\alpha^2}{n^2} = -\frac{2\pi^2 \epsilon^4 m}{n^2 \hbar^2}, \quad (144)$$

which is the Bohr energy.

111. *Root near $x=2$.* Put $x=2-\delta$. Then (137) gives

$$\delta = \frac{\alpha^2}{n^2} e^{-2} \left[1 + \frac{3}{2} \frac{\alpha^2}{n^2} e^{-2} + \dots \right]. \quad (145)$$

The leading terms give for the corresponding radius

$$r = \frac{\epsilon^2}{mc^2} \frac{1}{2-\delta} \sim \frac{1}{2} \frac{\epsilon^2}{mc^2} \left[1 + \frac{1}{2} \frac{\alpha^2}{n^2} e^{-2} \right]. \quad (146)$$

Since α^2 is very small, for different integer values of n these values of the radius all cluster round $\frac{1}{2}\epsilon^2/mc^2$.

112. This system, of two opposite charges ϵ , one associated with a proton and one with an electron, of the very small radius $\frac{1}{2}\epsilon^2/mc^2$, may be provisionally identified with the *neutron*. We then have the remarkable result that the different excited states of the neutron have all practically the same radius, namely, 1.4×10^{-13} cm. We shall see shortly that these orbits are unstable, and that unlike the complete set of Bohr orbits the only neutron orbits are these circular ones.

The value of the constant W for the neutron orbits is found to be

$$W \sim mc^2 \left(2 \frac{n}{\alpha} e^{-1} - 1 \right), \dots \dots \dots (147)$$

and is positive. The value of W' is found to be

$$W' \sim -mc^2 \left(2 \frac{n}{\alpha} e^{+1} + 1 \right), \dots \dots \dots (148)$$

and is negative.

From its definition, W' is the actual energy associated with the system in its state " n "; it is the sum of the kinetic and electric energies, as we calculate them in the present theory. On the other hand, W is equal to the energy of motion of the electron if the system disintegrated and the electron escaped to infinity. For by (129), putting $r = \infty$, $v = v_\infty$, $x = 0$, we have

$$W = mc^2 \left[\frac{1}{(1 - v_\infty^2/c^2)^{\frac{1}{2}}} - 1 \right] = T_\infty; \dots \dots \dots (149)$$

where T_∞ is the kinetic energy of the escaped electron at infinity.

This value of W , and so of T_∞ , is essentially positive. On the other hand, the actual energy of the neutron as constituted, W' , is essentially negative. It follows that when a neutron disintegrates the electromagnetic interactions between proton and electron result in the *generation* of an amount of energy $W + (-W')$, equal to $2mc^2(n/\alpha)(e^{+1} + e^{-1})$.

This generation of energy is something so different from our ordinary dynamical notions that we must be careful about our phraseology. The "binding energy" of the system is negative, as for any atomic system. But instead of our having to add energy to it to liberate an electron at infinity, it produces by its own disintegration an electron at infinity with the positive energy W . That fundamental rock of physics, the conservation of energy, is upset when a neutron disintegrates. For $n=1$, the binding energy is of the order of $-740 mc^2$ ergs or -370 million electron-volts; the energy W or T_∞ is of the order of $+100 mc^2$ ergs, or 50 million electron-volts.

Conversely, if a neutron is formed by the reverse process, by an electron at infinity with just the right energy spiralling into a circular orbit, then energy of amount T_∞ disappears, and, in addition, we would calculate that energy W' comes into existence. Since W' is negative, this means that further energy (W') has disappeared.

But we must be careful not to be entrapped by our own phraseology. The *formation* of a neutron by the reverse process is a highly improbable event. How a neutron comes into existence we do not know. But it is at least tenable, on these classical ideas, that it comes into existence through the liberating of an electron *very close to* a proton, with resulting capture. However this may be, as soon as we have a violation of the conservation of energy we are no longer entitled to say that the formation of a neutron represents the disappearance of a positive amount of energy $-W'$, just because W' is the negative energy of the system. All we can assert from our analysis is that kinetic energy W comes into existence as the energy of a high-speed electron every time a neutron disintegrates.

113. In the interiors of stars, on my theory of stellar structure, very high densities occur in the vicinity of the centre, and these conditions will be favourable for the formation of neutron orbits as temporary structures. Moreover, neutron-formation would seem to be an essential step in the synthesis of elements of high atomic numbers in stellar interiors. If, in fact, neutrons are thus formed near the centres of stars, their disintegration would be a powerful source of energy, and thus a possible source of *stellar* energy. The process would be capable of indefinite repetition, and so the centre region of a star may be considered as a system in a steady state yielding an inexhaustible supply of energy—in agreement with ideas put forward some years ago by Bohr and Landau.

114. We leave these speculations to investigate the orbits in general, for our system of proton and electron. By this means we shall be able to show that the circular orbits of the neutron are unstable.

We write the differential equation of the orbits in the form, from (133),

$$\frac{n^2 \left(\frac{dx}{d\theta} \right)^2}{\alpha^2} = \left(1 + \frac{W}{mc^2} \right)^2 e^x - e^{-x} - \frac{n^2}{\alpha^2} x^2. \quad (150)$$

Put $y = y(x) = \left(1 + \frac{W}{mc^2} \right)^2 e^x - e^{-x} - \frac{n^2}{\alpha^2} x^2. \quad (151)$

Then the integrated equation of the orbits is

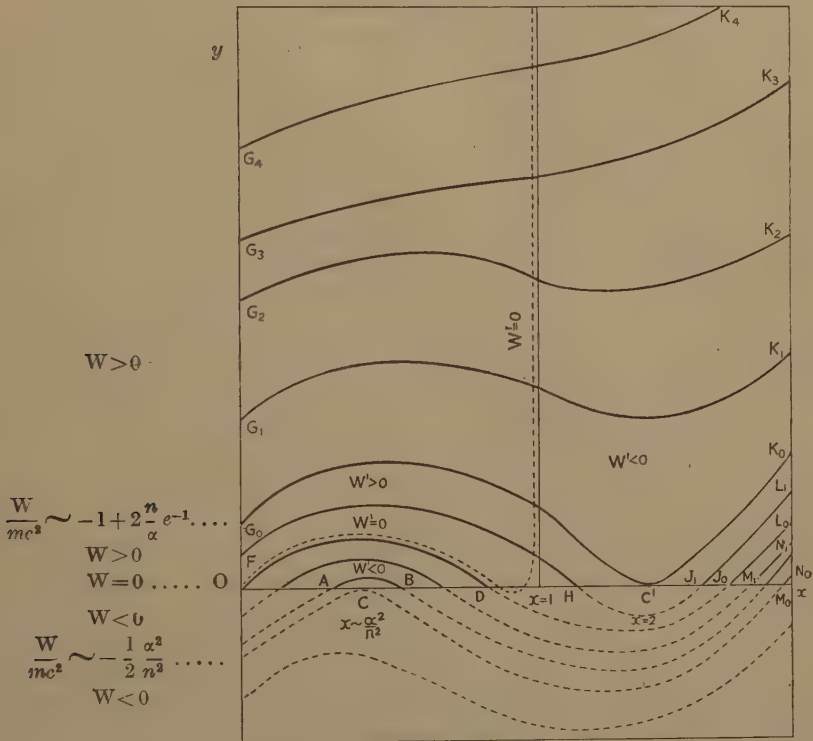
$$\theta + \text{const.} = \int \frac{(n/\alpha) dx}{y^{\frac{1}{2}}}. \quad (152)$$

115. By well-known theory the forms of these orbits are determined by the positive zeros of $y(x)$. These depend, in a somewhat complicated way, on the values of W and n . The best way of exhibiting the positions of the zeros is to plot the curves $y = y(x)$ for constant n , for a variety of values of W . It is only necessary to consider x positive. It would, however, be tedious to plot these curves accurately for specific numerical values of W . Moreover, such curves would be uninformative, since owing to the small value of α , the interesting and important features of the curves would be compressed into a microscopically small region of x .

Instead, we plot the curves $y=y(x)$ qualitatively, on a non-uniform scale in x ; we enormously exaggerate the scale of x for x small. Such a diagram is shown in fig. 1.

116. The general principle of the interpretation of these curves is that real orbits correspond to any portion of a curve $y=y(x)$ for which $y>0$. Apses correspond to the zeros of $y(x)$. If $y>0$ between two consecutive zeros of $y(x)$, then the corresponding orbit consists of "librations" between these two zeros, which represent inverse apsidal distances.

Fig. 1.



(Scale of x is non-uniform.)

Curves of constant W . The loci are :

$$y = (1 + W/mc^2)^2 e^x - e^{-x} - \left(\frac{n}{a}\right)^2 x^2.$$

The curve $W'=0$ is also shown, dotted. Its equation is

$$y = e^{-x} \left[\frac{1}{(1-x)^2} - 1 \right] - \left(\frac{n}{a}\right)^2 x^2.$$

If $y>0$ between $x=0$ and the first positive zero of $y(x)$, then the corresponding orbit passes from $r=\infty$ to the corresponding apse and back again. If $y>0$ between $x=\infty$ and the last positive zero of $y(x)$, then the corresponding orbit passes from the nucleus outwards, reaches the

apsidal distance corresponding to the zero in question, and returns to the nucleus. For a value of W for which $y(x)$ has no zero and remains positive for $x > 0$, the corresponding orbit extends from the nucleus to infinity.

117. It will be seen from fig. 1 that, according to the values of W (for given n), the curve $y=y(x)$ may have one, two or three positive zeros; and two of these may coincide, as we have already seen from our discussion of circular orbits. It is convenient to begin our enumeration of the various possibilities with that curve $y=y(x)$ which touches the x -axis at C , the point corresponding to the Bohr circular orbits; C is the point $x \sim \alpha^2/n^2$, and for this curve $W \sim -\frac{1}{2}\alpha^2/n^2$. This curve cuts the x -axis again in a point beyond $x=2$. It is wholly below the x -axis up to this distinct zero (M_0 in fig. 1), wholly above the x -axis beyond it. Curves for slightly different values of W either meet the x -axis in two points near C or do not meet it at all near C . For values of W between zero and $-\frac{1}{2}\alpha^2/n^2$, $y=y(x)$ meets the x -axis in two points, such as A , B , between which $y > 0$; accordingly the corresponding orbit consists of librations between the apses corresponding to A and B . For values $W < -\frac{1}{2}\alpha^2/n^2$, $y=y(x)$ does not meet the x -axis near C . It follows from all these considerations that the Bohr circular orbits are *stable*; for disturbance to a neighbouring value of W can at most change the circular orbit into a neighbouring orbit with apses corresponding to A and B .

It follows also that, in a Bohr orbit of varying r or x , although the energy W' is not constant during the libration, there is no secular decrease of W' ; the value of W' fluctuates; but there can be no radiation of energy. In the present example Bohr orbits are strictly radiationless, in spite of the acceleration of the electron.

The complete set of Bohr non-circular orbits for given n corresponds to a sub-set of the arcs such as AB . The last of these is the arc OD , which passes from "infinity" at O to an apse at D , and corresponds to $W=0$, and very closely also to $W'=0$.

Each of these curves corresponding to a Bohr orbit meets the x -axis again in points such as J_0 , M_1 , M_0 , beyond which they rise once more above the x -axis. The arcs extending beyond J_0 , M_1 , M_0 , thus correspond to orbits of negative W , "bound" to the nucleus, having apses at J_0 , M_1 , M_0 , . . . , and passing from the nucleus outwards to these apses and then returning to the nucleus.

118. For small positive values of W we get arcs such as FH , which correspond to a positive value of the energy at "infinity," with an apse at H . The last of this family touches the x -axis at C' , near $x \sim 2$, and passes beyond this to the nucleus ($x = \infty$). In the vicinity of C' the curves for $y > 0$, which alone correspond to real orbits, do not meet the x -axis in real points, and so a circular orbit, corresponding to the point C' itself, is unstable; for if the orbit is disturbed to a neighbouring value of W , the moving particle must pass either to the nucleus, along the

arc $C'K_0$, or to freedom, along the arc $C'G_0$. These orbits are orbits of positive W . For still larger values of W , the curves $y=y(x)$ do not meet the positive x -axis in real points, and the corresponding orbits pass from infinity to the nucleus.

119. It is scarcely necessary to mention that "infinity" throughout this discussion is used in a purely conventional sense; the orbits are necessarily always contained in the substratum itself, which though finite on the t -scale is infinite on the τ -scale.

120. Although the circular orbits in the neighbourhood of C' are unstable, the orbit G_0C' which touches the x -axis at C' describes an infinite number of convolutions round the nucleus as it approaches C' . The circular orbit is, in fact, a circular asymptote to the orbit GC' , and consequently a particle describing G_0C' spends an infinite amount of time, theoretically, in the vicinity of the circular orbit. Similarly, the orbit K_0C' has the circular orbit C' for an asymptotic circle, and spends an infinite amount of time spiralling into it. Orbits which pass from the nucleus to infinity, or in the reverse direction, by going close to C' without actually passing through C' spend a large but finite time in the vicinity of the circular orbit. It is apparent from these considerations that, though the circular orbits at C' are unstable, the system is capable of a transient existence in this configuration. Hence neutrons should "exist" for short though finite times.

121. Let us calculate the time-relations along an orbit. From the integrals of motion (129) and (130) we have

$$mr^2 \frac{d\theta}{d\tau} = H \left(1 + \frac{W}{mc^2} \right)^{-1} e^{-\frac{1}{2}x},$$

whence the time to a "point x " is given by

$$\tau = \int^\tau d\tau = \frac{m}{H} \int^x \left(1 + \frac{W}{mc^2} \right) e^{\frac{1}{2}x} r^2 d\theta.$$

Substituting for $d\theta$ from (150) and replacing r in terms of x , we find

$$\begin{aligned} \tau &= \frac{\epsilon^2}{mc^3} \int^x \frac{(1+W/mc^2)e^{\frac{1}{2}x} dx}{x^2 y^{\frac{1}{2}}} \\ &= \frac{\epsilon^2}{mc^3} \int^x \frac{(1+W/mc^2) dx}{x^2 [(1+W/mc^2)^2 - e^{-2x} - (n^2/\alpha^2)x^2 e^{-x}]^{\frac{1}{2}}}. \quad (153) \end{aligned}$$

This integral *diverges* near a double root of $y=0$, and hence, along an orbit which spirals into the circular orbit, the moving particle spends an infinite amount of time in the vicinity of the circular orbit. The integral *converges* at $x=\infty$, and hence a particle takes a finite time to reach the nucleus along an orbit of the type of G_1K_1 or J_1L_1 (fig. 1). The order of magnitude of this time is ϵ^2/mc^3 , which shows at a glance that the order of magnitude of the velocity along these orbits in the vicinity of the nucleus is equal to c . An orbit which has *very nearly* a double

zero of y along its course will yield a large contribution to (153), so that an orbit passing close to the circular orbit will spend a large time in this vicinity.

122. It will be now apparent that the circular orbits near $x \sim 2$ which we have provisionally identified with the neutron have a much wider significance than their instability would appear to connote. The neutron radius $x \sim 2$ acts substantially as a *barrier* to particles proceeding outwards from the nucleus, or to particles proceeding inwards from infinity. An electron attempting to escape from the nucleus is either held up by this barrier and compelled to return to the nucleus (orbits such as J_1L_1 , M_1N_1 , etc.), or just surmounts the barrier, but take an enormous time to do so. Similarly, an incoming electron behaves as though it encounters a barrier in the same position. With insufficient energy it is turned back and retreats to infinity again. With energy near the critical value, it spends a long time in the vicinity of the barrier. With enough energy it surmounts the barrier, though spending a long time in its vicinity, and finally reaches the nucleus. There is, of course, no actual potential barrier; the law of force has remained throughout the Coulomb law; but the dynamics we have had no choice but to adopt has had consequences closely similar to the known and well-recognized feature of the existence of a potential barrier in the neighbourhood of the frontier of the nucleus.

123. The diagram (fig. 1) exhibits the curves of constant W . It would be possible to graph on the same diagram curves of constant *energy* W' . These coincide extremely closely with the curves of constant W for x very small, *i. e.*, inside the Bohr region ODCO, but outside this they differ widely. It is sufficient to consider the boundary curve $W'=0$ which divides regions of positive binding energy from regions of negative binding energy. It is shown as a dotted curve in fig. 1 (not to be confused with the dotted portions of the curves $W=\text{const.}$ in the lower half of the figure, which correspond to y negative). This curve $W'=0$ starts at O and first follows closely, but just outside the curve OD ($W=0$), which corresponds to the pseudo-parabolic Keplerian orbit. After D it crosses and recrosses the axis of x , the first crossing being in the neighbourhood of $x=2(\alpha^2/n^2)$ and the next just before $x=1$; it then ascends steeply, becoming asymptotic to $x=1$. This curve W' shows that, inside the arc OD, W' is negative; above OD, W' is positive; and to the right of $x=1$, W' is negative again. Thus all nuclear-bound orbits, as well as all the Bohr orbits, possess negative energies W' . Any completely penetrating orbit such as G_1K_1 passes from W' positive to W' negative.

124. The energy W' is not strictly constant over a Bohr orbit, but oscillates slightly between values corresponding to the two extreme values of x (apsidal distances) for the orbit in question. This may have a bearing on the natural width of spectral lines: if the system is disturbed, it may be supposed to pass from one value of n to another,

but the actual values of W' between which it makes a transition will not be fixed, but depend on the part of the orbit in which the electron happens to be.

125. It will be apparent that though we have in this paper explicitly confined our attention to the orbits of a single electron, of charge $-\epsilon$, in the presence of a single fixed proton, of charge $+\epsilon$, our results have a much wider significance. All our formulæ are at once generalized to a nucleus of atomic number Z , on replacing ϵ^2/mc^2 by $Z\epsilon^2/mc^2$, and α by $Z\alpha$. Thus, for example, such a nucleus will exhibit the features of an apparent potential barrier. Previously this has been an *ad hoc* assumption, which fructified especially in the hands of Gamow; in our work it is a deductive consequence of our rational electrodynamics.

126. Again, we must suppose that a nucleus of atomic number Z has itself a complicated structure. An electron "bound" to it along orbits such as J_1L_1 , M_1N_1 may suffer considerable perturbation as it gets deeper into the nuclear structure—we are not compelled to assume that it continues to move in towards the nucleus and finally reaches the point $x=\infty$, $r=0$, as we did for the neutron. But the new point is that these perturbations may transfer it from an orbit of the type N_1M_1 , or L_1J_1 , to an orbit of the type K_1G_1 , which is capable of leaving the nucleus altogether, although of negative W' . Inside the "nuclear region" there is continuity between these two types of orbit, and when there is a probability of an electron being transferred from the nuclear-bound types of orbit to the escaping type of orbit, the nucleus in question will appear as β -radio-active, and will violate the conservation of energy. The minimum emergent energy for which this can happen will be of the order of $(2/Z\alpha)(2.72)^{-1}mc^2$, as shown by the value of W for the curve of type $G_0C'K_0$. For Z of the order of 100, this energy is of the order of mc^2 ergs or 5×10^5 electron-volts, which is of the observed order of magnitude of nuclear β -ray activity.

127. On the other hand, electrons bound to the "nucleus" will be in general unable to move beyond a distance of the order of $Z\epsilon^2/mc^2$, where they will be stopped by the apparent potential barrier. It may be noticed that for Z small, very large accession of energy W' would be required to transfer an electron from the bound type of orbit such as J_1L_1 to the escaping type of orbit such as K_1G_1 , and hence, although conservation of energy is not either assumed or obeyed, we may expect that such transitions will be unlikely for small Z . Hence β -ray activity for elements of low atomic number is unlikely. Since conservation of energy does not hold for these high-speed electrons moving in such close proximity to a positive charge, there is no meaning in asking where the energy comes from, in the case of the expulsion of an electron; it does not come, for example, from some pre-existing store of energy. The energy is manufactured as the electron is expelled, by the electromagnetic forces. The upper limit to the continuous β -ray spectrum

observed experimentally may be presumed to be due to some condition imposing a maximum to alterations in the constant W , or the energy W' , in intra-nuclear transitions of the type under discussion.

128. The further investigation of the possibilities of interaction of particles of differing signs on the present rational electrodynamics offers an immense field for research. It is not to be supposed that all the details of the present dynamics will prove to be correct; the main thing is that we have constructed a new, relativistic, non-Newtonian, non-Einsteinian dynamics, and a corresponding electrodynamics which exhibits in the particular cases here studied, in spite of strict Coulomb interactions, the features of an apparent radius of the electron, an apparent breakdown of the Coulomb law at this radius, intra-nuclear bound orbits of negative energy W' but positive integration constant W , non-conservation of energy and the possibility of β -radio-activity. We have the beginnings of a true nuclear dynamics, based solely on the development of rational electrodynamics, and free from all specific assumptions. Much further work remains to be done, even in the domain of the Bohr orbits, in investigating the consequences of the integrals of "energy" and "momentum," (129) and (130); for example, we could trace the size of the variations from pure Keplerian orbits. Further, a whole statistical mechanics needs to be constructed, for handling large numbers of particles in a system such as a nucleus of large atomic number. There is, in fact, a big "pseudo-classical" field waiting to be explored, before quantum considerations are superposed. Just as our introduction, in Paper III. of this series, of a term in the magnetic effect of a moving charged particle on a moving test-particle, depending on the velocity of the latter, largely obviates the need for the special assumption of "electron spin," and yielded the peculiar exponential $e^{\frac{1}{2}x}$ in (130), so it is to be expected that a number of phenomena, hitherto attributed to quantum effects, may be explained on classical grounds. I am aware that not all readers will agree with this view; but I have at least shown that the field of "classical" methods is capable of enormous extension.

129. I conclude by briefly considering the interaction of particles of like sign. We now put $e_1 = +e$, $e_2 = +Ze$, and write (choosing them to be protons)

$$x' = \frac{Ze^2}{m_H c^2} \frac{1}{r}.$$

The integrals of "energy" and "momentum" now take the form

$$\frac{1}{(1-v^2/c^2)^{\frac{1}{2}}} = \left(1 + \frac{W}{m_H c^2}\right) e^{-x'}, \quad . \quad . \quad . \quad . \quad (154)$$

$$\frac{m_H v^2 d\theta/d\tau}{(1-v^2/c^2)^{\frac{1}{2}}} = H e^{-\frac{1}{2}x'}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (155)$$

Eliminating $d\tau$ between these equations as before, we find that the differential equation of the orbit takes the form

$$x'^2 + \left(\frac{dx'}{d\theta}\right)^2 = \frac{Z^2\epsilon^4}{H^2c^2} = \left[\left(1 + \frac{W}{m_Hc^2}\right)^2 e^{-x'} - e^{x'} \right]. \quad (156)$$

Putting $dx'/d\theta=0$, the apses are given by

$$Y(x')=0, \quad (157)$$

where
$$Y(x') \equiv \left(1 + \frac{W}{m_Hc^2}\right)^2 e^{-x'} - e^{x'} - \frac{H^2c^2}{Z^2\epsilon^4} x'^2. \quad (158)$$

The function $Y(x')$ is utterly unlike the function $y(x)$ we studied before; $Y(x')$ steadily decreases as x' increases from zero, and there is a single zero for any W, H . Consequently there is a single apse on all orbits, at perihelion, and all orbits return from this apse both ways to infinity. We may, therefore, most simply calculate W and H in terms of the motion at infinity. If V is the speed at infinity, d the distance of the asymptote from the nucleus, the constants W, H are given by putting $x'=0$ in (154) and (155), namely,

$$\frac{1}{(1-V^2/c^2)^{\frac{1}{2}}} = 1 + \frac{W}{m_Hc^2}, \quad (159)$$

$$\frac{m_H V d}{(1-V^2/c^2)^{\frac{1}{2}}} = H. \quad (160)$$

Hence (156) becomes

$$x'^2 + \left(\frac{dx'}{d\theta}\right)^2 = \frac{Z^2\epsilon^4}{m_H^2 V^2 d^2 c^2} \left[e^{-x'} - \left(1 - \frac{V^2}{c^2}\right) e^{x'} \right]. \quad (161)$$

The distance of the apse for a head-on collision is given by solving this equation for x' when $(dx'/d\theta)=0$ and d is put equal to zero. This gives

$$x' = \frac{1}{2} \log \frac{1}{1-V^2/c^2}, \quad (162)$$

i. e., the apsidal distance is

$$\frac{Z\epsilon^2}{m_Hc^2} \frac{2}{[\log (1-V^2/c^2)]}. \quad (162')$$

Even for V quite close to c , the order of magnitude of this is but little removed from $2Z\epsilon^2/m_Hc^2$, which is thus the order of magnitude of the distance of closest approach of a proton and a nucleus of atomic number Z .

130. Equation (161) determines the angle of scattering of the colliding-proton. Its deflexion is the angle 2Θ , where

$$\Theta = \int_0^{\infty} \frac{dx'}{\left[\frac{Z^2\epsilon^4}{m_H^2 V^2 d^2 c^2} \left\{ e^{-x'} - \left(1 - \frac{V^2}{c^2}\right) e^{x'} \right\} - x'^2 \right]^{\frac{1}{2}}} \quad (163)$$

and a is the zero of the denominator. It would be interesting to use this formula to derive the proportions of protons scattered through various angles, in the usual way, and to compare it with experimental results. This must be left over for a future occasion. Here it is sufficient to mention that (163) will depart widely from the Rutherford formula when a is not small, i. e., when the apsidal distance is of the order of $Z\epsilon^2/m_H c^2$.

131. *Summary of Paper V.*—The rational electrodynamics developed in the preceding series of papers is applied to the case of an electron, of charge $-\epsilon$ and mass m , interacting with a proton, of charge $+\epsilon$ and mass m_H . It is shown that when the proper relativistic equations of motion are integrated, though they possess constants of motion W and H , W and H are not precisely equal to the energy and angular momentum respectively. The latter are in consequence no longer constant.

The integrals are applied to yield the orbits. The remarkable fact emerges that, in addition to a set of orbits corresponding very closely to the Keplerian orbits, a new set of orbits appears. When these are quantised in H , it is found that in addition to the Bohr orbits, there exist circular orbits of radius $\frac{1}{2}\epsilon^2/mc^2$, which give systems corresponding to the existence of the *neutron*. For different quantum numbers n , the different neutron radii are closely coincident. These neutron circular orbits are *unstable*, but act as asymptotic circular orbits for particles approaching the nucleus from outside, or proceeding from the nucleus outwards, for a certain value of W ; and, in consequence, though the circular orbits are *unstable*, an electron will in general spend a long time in their vicinity before either escaping to infinity or falling into the nucleus. It follows, further, that the proton appears to possess a *potential barrier* at the distance $\frac{1}{2}\epsilon^2/mc^2$, though this is a dynamical phenomenon, and not due to any “break-down” of the Coulomb law.

The non-circular orbits are analysed in detail. It is shown that the *conservation of energy is violated* when an electron passes through the apparent potential barrier, a positive energy at infinity being derived from a negative energy for the same electron, in the same orbit, when near the nucleus. Thus *energy* appears to be *generated* when a neutron disintegrates. In consequence of the violation of the conservation of energy and of angular momentum, great care must be exercised in the use of ordinary dynamical phraseology. The ideas derived from this study of the strict “classical” relativistic interaction of electron and proton are briefly applied to nuclear dynamics in general, and shown to be compatible with β -radioactivity

XXIX. *Voltage Fluctuations in Resistors. (Johnson Effect.)*

By N. R. CAMPBELL, Sc.D., F.Inst.P., and
V. J. FRANCIS, B.Sc., A.R.C.S., F.Inst.P.*

(Communication from the Research Laboratories of The General Electric Company Limited, England.)

[Received October 16, 1942.]

SUMMARY.

Gives a proof of Johnson's formula for thermal fluctuations based on the principles of Brownian motion and exhibiting the relation between the Johnson effect in resistors and the shot effect in a saturated electronic current.

1. JOHNSON (1928) gave the following formula for the fluctuations that appear in valve circuits owing to the thermal energy of the resistors forming part of them :

$$\overline{I^2} = \frac{2kT}{\pi} \int_0^\infty R(\omega) |Y(j\omega)|^2 d\omega. \quad (1.1)$$

Here it is supposed that an impedor, whose impedance has the real part $R(\omega)$, is connected across the input (otherwise of infinite impedance) of a valve-amplifier whose transfer-admittance is $Y(j\omega)$. T is the absolute temperature of the impedor and k is Boltzmann's constant. I is the fluctuating current in the output circuit and $\overline{I^2}$ its mean square deviation from its mean, which is zero.

Johnson also pointed out that, if $Y(j\omega)$ is non-zero only within some range $\delta\omega$ and has a constant value within this range, and if $R(\omega)$ has the constant value R within the range, then $\overline{I^2}$ has the same value as it would have if the impedor had in series with it, or were replaced by, a source of sinusoidal E.M.F. having any frequency within the range, and having a mean square value

$$\overline{E^2} = 2kTR \delta\omega / \pi = 4kTR df. \quad (1.2)$$

Johnson supported his formula by experiments only. Nyquist (1928) professed to deduce the formula from the principles underlying the Rayleigh-Jeans law of thermal radiation ; Burgess (1941) gave another proof based on the same principles. Bernamont (1937), Bakker and Heller (1939), and Spenke (1939) have deduced the formula from the electronic theory of conduction and thus related the effect to Brownian motion rather than to thermal radiation.

* Communicated by C. C. Paterson, F.R.S.

2. The object of this paper is to show that (1.1) can be deduced from the principles of Brownian motion without making any assumption about the mechanism of conduction. The fundamental assumption, based on those principles, is that the fluctuations are due to a large number of events happening randomly with respect to time in the impedor or (since the effect is known to be associated with resistors but not with reactors) in the resistors forming part of it. The Johnson effect is thus brought into close relation with the shot effect, which is also due to events random in time; the only difference between them is in the nature and location of the events.

The physical assumptions concerned in the proof will be discussed fully, but many of the mathematical propositions involved will be stated baldly without proof. They are all straightforward applications of the operational calculus and the theory of Fourier transforms, similar to those involved in our recent discussion of the shot effect (Campbell and Francis, 1941). Quotations from that discussion will be denoted by the letter A, so that (A 11.6) means equation (11.6) of that paper.

3. We have first to make an assumption about the nature of the events in the resistor, that is to say (A, § 12) we require a correspondence proposition between an atomic event in a resistor and some large scale event accessible to experiment. In order that they may affect the output of the amplifier, the events must produce voltage changes at the input. Accordingly we shall assume that each event in the resistor is equivalent, in respect of any circuit of which the resistor forms part, to the generation of a momentary E.M.F. This may appear at first sight not to be the most plausible assumption about the nature of the events. For in a metallic resistor the events are undoubtedly the passages of the electrons along their free path, and such a passage may seem equivalent to a current rather than an E.M.F. But the equivalent current is not strictly independent of the circuit connected to the resistor; Spenke (*l. c.*) has pointed out that calculations, based on the equivalence of free flights to currents, are valid strictly only when the resistor is short circuited externally. In order to attain results applicable in other circumstances, assumptions have to be introduced which, whatever their form, make a free flight equivalent to the generation of a definite E.M.F. independent of the external circuit.

In order that resistors in parallel should be equivalent to a single resistor according to the usual laws, it is necessary that the source of the E.M.F. should be in series, and not in parallel, with the resistor. We may then regard all the momentary E.M.F.'s as located in an impedanceless generator in series with the resistor.

Our assumption, stated more completely, is that there is in series with each resistor an impedanceless generator, and that the chance that an E.M.F., lasting for a period L , and having during that period the mean value e , shall start during the interval dt is $\lambda_e dt$. There may be an indefinite number of different kinds of E.M.F.'s characterized by different

suffices r . Since the mean voltage between the ends of the resistor is zero, we must have

$$\Sigma_r \lambda_r e_r L_r = 0.$$

4. It is convenient to start with the further assumptions that the impedor across the input is a single resistor of resistance R , and that L_r is short compared with any frequency to which the amplifier responds; more definitely that $Y(\omega) = 0$ for any ω comparable with any $1/L_r$. Then, if $I = s(t)$ represents the current in the output circuit due to a single event of the r th kind, it follows from A (11.6) that the Fourier transform of $s(t)$ is

$$S(j\omega) = \frac{e_r L_r}{\sqrt{2\pi}} Y(j\omega). \quad (4.1)$$

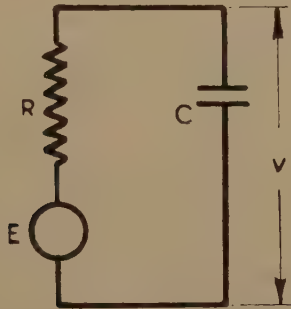
Then, applying a generalized form of A (4.4)

$$\overline{I^2} = \frac{\Sigma_r \lambda_r e_r^2 L_r^2}{\pi} \int_0^\infty |Y(j\omega)|^2 d\omega. \quad (4.2)$$

If this is to agree with (1.1) we must have

$$\Sigma_r \lambda_r e_r^2 L_r^2 = 2kTR. \quad (4.3)$$

Fig. 1.



The work of Bernamont, etc. (*l. c.*), is in effect a proof that this is true when the theory of electronic conduction applies to the resistor. We require a wider basis for (4.3). It may be found in the well-known proposition that the fluctuations of voltage v across a capacitor C in parallel with a resistor, the combination being at a temperature T , is given by

$$\overline{v^2} = kT. \quad (4.4)$$

This proposition results directly from the doctrine of the equipartition of energy and the fact that such a combination has one degree of freedom.

In such a combination (fig. 1) the transfer-ratio between the hypothetical generator E , in series with the resistor and the terminals of the capacitor, is $1/(1 + j\omega CR)$. Accordingly, in analogy with (4.2),

$$\begin{aligned} \overline{v^2} &= \frac{\Sigma_r \lambda_r e_r^2 L_r^2}{\pi} \int_0^\infty \left| \frac{1}{1 + j\omega CR} \right|^2 d\omega \\ &= \Sigma_r \lambda_r e_r^2 L_r^2 \cdot \frac{1}{2RC}. \end{aligned} \quad (4.5)$$

If (4.5) is to agree with (4.4), (4.3) must be true.

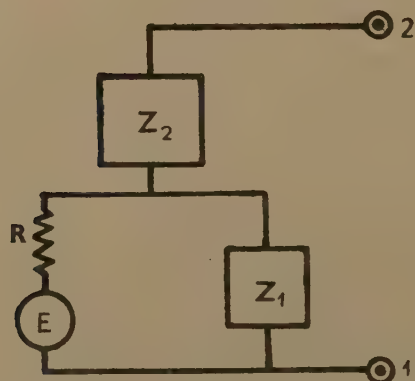
5. We now remove one of the restrictions of § 4, and suppose that the elements connected across the input of the amplifier include reactors as well as one or more resistors. Let us first suppose that there is only one resistor, the other elements being all reactors. The presence of reactors will make the voltage appearing between the input terminals, in virtue of a single event, different from the E.M.F. generated in the resistor; even if the E.M.F. is an instantaneous pulse, the consequent input voltage will in general have a finite duration. Let the input voltage corresponding to an event of the r th kind be represented by

$$v_r = e_r L_r \xi(t). \quad \dots \dots \dots (5.1)$$

If $\mathcal{E}(j\omega)$ is the Fourier transform of $\xi(t)$, we have in place of (4.2) in analogy with (A 5.5)

$$\overline{I^2} = \frac{\Sigma \lambda_r e_r^2 L_r^2}{\pi} \int_0^\infty 2\pi |\mathcal{E}(j\omega) Y(j\omega)|^2 d\omega. \quad \dots \dots \dots (5.2)$$

Fig. 2.



If this is to agree with (1.1) we must have

$$\Sigma \lambda_r e_r^2 L_r^2 2\pi |\mathcal{E}(j\omega)|^2 = 2kTR(\omega). \quad \dots \dots \dots (5.3)$$

This is clearly possible, since both $\mathcal{E}(j\omega)$ and $R(\omega)$ are determined by the nature of the elements; it can be proved to be true in simple cases by direct calculation. Thus we may consider the combination of a resistor R with two pure reactors, Z_1 in parallel and Z_2 in series (fig. 2.) The transfer-ratio between the hypothetical generator E and the terminals 1, 2 of the whole combination is $Z_1/(R+Z_1)$; so long as the impedance across (1, 2) is infinite, Z_2 does not affect the ratio. Hence by another application of (A 11.6), remembering that $Z_1(j\omega)$ is wholly imaginary,

$$\mathcal{E}(j\omega) = \frac{1}{\sqrt{2\pi}} \frac{Z_1(j\omega)}{R+Z_1(j\omega)}. \quad \dots \dots \dots (5.4)$$

$$2\pi |\mathcal{E}(j\omega)|^2 = \frac{|Z_1(j\omega)|^2}{R^2 + |Z_1(j\omega)|^2}. \quad \dots \dots \dots (5.5)$$

On the other hand,

$$R(\omega) = \text{real part of } \left\{ Z_2(j\omega) + \frac{R \cdot Z_1(j\omega)}{R + Z_1(j\omega)} \right\} \quad (5.6)$$

Since $Z_2(j\omega)$ is also imaginary

$$R(\omega) = \frac{R Z_1 |j\omega|^2}{R^2 + |Z_1(j\omega)|^2}, \quad (5.7)$$

so that

$$R \cdot 2\pi |E(j\omega)|^2 = R(\omega). \quad (5.8)$$

But (5.8) is the relation that must hold if (5.3) is to follow from (4.3) and (5.2).

6. When there are several resistors among the elements across the input, we should naturally proceed by dividing the whole combination into several parts, each containing only one resistor. But now these parts are not each across an input of infinite impedance; some will be shunted by others. It is not immediately clear that a knowledge of the change in the form of the pulse due to the presence of the reactors, calculated when each part of the combination is connected in turn across the input, will tell us how the combination as a whole behaves.

However, the natural procedure is made possible by the following proposition, which is quite general. If the open circuit voltage across any two-terminal network, consisting of sources of E.M.F. and impedors is v , and if the impedance of the network is Z , then, if the network forms part of any more complicated network, it is equivalent to a source of E.M.F. $e=v$ in series with an impedor of impedance Z . Accordingly we may calculate the voltage across the terminals of any one of the parts when that part alone is across the input and also the impedance of the part. Then, when the part is combined with others, it will behave as a generator of E.M.F. equal to the calculated voltage in series with the calculated impedance.

We may now proceed in one of two ways. Following Bell (1939), we may consider two resistive combinations connected in series by impedanceless connexions in a constant temperature enclosure; one is a pure resistor, the other a combination of resistors and reactors. Then, it is argued, the average power dissipated in the reactive impedor by the pulses arising in the generator in series with the resistor must be equal to the average power dissipated in the resistor by the pulses arising in the generator in series with the reactive impedor. The argument actually leads to (5.3) in virtue of the fact that the power dissipated in an impedor depends only on the real part of the impedance. It has the merit that it is independent of the nature of the impedor, so long as the impedor has terminals in series with which the generator may be placed; thus it does not assume that the reactance is lumped. It has the demerit of all very general and sweeping arguments of this kind, namely the difficulty of establishing, in view of possible unsuspected complications, that the supposed equilibrium actually exists. The argument gains strength if it can be confirmed by more particular arguments; accordingly we shall show that, at least in the case of lumped impedances, (5.3) is a consequence of direct circuit analysis.

Let three impedors be arranged as in fig. 3 across the input of an amplifier of infinite input impedance. Let $R(\omega) + jX(\omega)$ with suffixes 1, 2, 3 be the impedances of the three impedors, both $R(\omega)$ and $X(\omega)$ being real. Suppose that impedor n is equivalent to a generator, in series with impedance $R_n(\omega) + jX_n(\omega)$, characterized by

$$\Sigma \lambda_r e_r^2 L_r^{-2} \cdot 2\pi |\mathcal{E}_n(j\omega)|^2 = 2kTR_n(\omega). \quad (6.1)$$

The transfer-ratio between the generator of impedor 1 and the output of the amplifier is

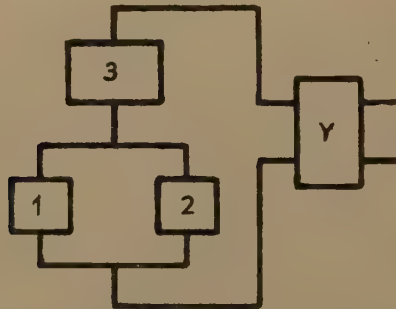
$$\frac{R_2(\omega) + jX_2(\omega)}{R_1(\omega) + R_2(\omega) + j\{X_1(\omega) + X_2(\omega)\}} Y(j\omega), \quad (6.2)$$

and the contribution of impedor 1 to the output is

$$\overline{v_1^2} = \frac{2kT}{\pi} \int_0^\infty R_1(\omega) \left| \frac{R_2(\omega) + jX_2(\omega)}{R_1(\omega) + R_2(\omega) + j\{X_1(\omega) + X_2(\omega)\}} Y(j\omega) \right|^2 d\omega. \quad (6.3)$$

The contribution $\overline{v_2^2}$ of impedor 2 is (6.3) with suffixes 1 and 2 inter-

Fig. 3.



changed. The transfer-ratio between the generator of impedor 3 and the output is $Y(j\omega)$. Hence the contribution of impedor 3 is

$$\overline{v_3^2} = \frac{2kT}{\pi} \int_0^\infty R_3(\omega) |Y(j\omega)|^2 d\omega. \quad (6.4)$$

Since the events are random in time, the whole contribution is

$$\overline{v^2} = \overline{v_1^2} + \overline{v_2^2} + \overline{v_3^2}. \quad (6.5)$$

Performing the necessary algebra, we find, dropping the argument ω for brevity,

$$\overline{v^2} = \frac{2kT}{\pi} \int_0^\infty \left\{ \frac{R_1(R_2^2 + X_2^2) + R_2(R_1^2 + X_1^2)}{(R_1 + R_2)^2 + (X_1 + X_2)^2} + R_3 \right\} Y^2 d\omega. \quad (6.6)$$

On the other hand, the real part of the impedance of the whole combination is, again dropping the argument on the right-hand side, the real part of

$$R_3 + jX_3 + \frac{\{R_1 + jX_1\}\{R_2 + jX_2\}}{R_1 + R_2 + jX_1 + jX_2} \quad (6.7)$$

$$= R_3 + \frac{(R_1 R_2 - X_1 X_2)(R_1 + R_2) + (R_1 X_2 + R_2 X_1)(X_1 + X_2)}{(R_1 + R_2)^2 + (X_1 + X_2)^2}, \quad (6.8)$$

which is equal to the expression in $\{ \}$ in (6.6).

We have therefore proved that, if each of the three impedors is equivalent to a generator characterized by (5.3) or (6.1) in series with its impedance, then a combination of these impedors arranged as in fig. 3 is also so equivalent. But, in virtue of our general circuit proposition and (5.5), (5.7), we have proved that one kind of impedor, namely that of fig. 2, is so equivalent. Any impedor whatever, made up of lumped impedances, can be built up by combining impedors of this kind according to fig. 3, combining the resulting impedors again according to fig. 3, and so on. (Of course in some of the impedors either R_n or X_n may be zero or infinite, subject to the condition that the impedance of the whole combination is never infinite.) It follows that (5.3) is true of any impedor of finite impedance, and therefore that (1.1) is true of any such impedor.

7. We have now to remove the other restriction of § 4, namely that $1/L_r$ is large compared with any frequency for which $Y(j\omega)$ is non-zero. That is to say, we have to suppose that the pulses are represented by (5.1) and that $\xi(t)$ is non-zero over a finite range of t . If $\xi(t)$ has any form appropriate in the sense of § 5 to some combination of lumped impedances, in other words, if an impedor which acts as a pure resistor at low frequencies always acts as a combination of lumped impedances at high frequencies, then the general validity of (1.1) may appear to follow from the preceding argument. This hypothesis is plausible; and Spence (*l. c.*) professes to have proved a proposition that seems to be equivalent. However, even if it is true, there is a difficulty. The whole argument is based on the assumption that the combination of a "resistor" in parallel with a capacitor has one degree of freedom. But a combination with a capacitor of an impedor having lumped impedances may have more than one degree: roughly the condition that it shall have only one degree is that the impedor shall contain no capacitor that cannot be merged in the parallel capacitor. We have not therefore proved (1.1) strictly for very high frequencies. But since L_r for an electronic conductor is about 10^{-14} sec., and is likely to be as little as 10^{-9} sec. for any kind of conductor, it is probably impossible to ascertain by experiment whether (1.1) is actually true when $1/L_r$ is comparable with frequencies to which the amplifier responds. The possibility must remain that it is not true in these conditions.

References.

- Bakker, C. J., and Heller, G., *Physica*, vi. p. 262 (1939).
 Bell, D. A., *Phil. Mag.* xxvii. p. 645 (1939).
 Bernamont, H. J., *Annales de Physique*, vii. p. 71 (1937).
 Burgess, R. E., *Proc. Phys. Soc. Lond.* liii. p. 293 (1941).
 Campbell, N. R., and Francis, V. J., *Phil. Mag.* xxxii. p. 496 (1941).
 Johnson, J. B., *Phys. Rev.* xxxii. p. 97 (1928).
 Nyquist, W., *Phys. Rev.* xxxii. p. 110 (1928).
 Spence, E., *Wiss. Ver. Siemens*, xviii. No. 2, p. 55 (1939).

XXX. *Aristotle, Newton, Einstein* *.

By E. T. WHITTAKER †.

[Received January 22, 1943.]

IT falls to us this year to commemorate the greatest of men of science, Isaac Newton, on the occasion of the three-hundredth anniversary of his birth. The centuries have not dimmed his fame, and the passage of time is unlikely ever to displace him from the supreme position. His discoveries, however—and this is part of their glory—have not persisted unchanged, but in the hands of his successors have been continually unfolding into fresh evolutions. During the eighteenth and nineteenth centuries there was an immense expansion of knowledge, springing directly from his work, and forming ultimately a vast superstructure based on the Newtonian concepts of space, mass, and force. Since 1900 the progress of science has continued, but the development of physics has changed in character: it has become subversive and radical, questioning the traditional assumptions and uprooting the old foundations. In 1915 the Newtonian doctrine of gravitation was superseded by that of Einstein: the divergence between the results of the two theories, so far as concerns the calculation of the movements of the planets, is extremely slight, and indeed, in almost all cases, too small to be detected by observation: but on the question of the essential nature of gravitation, the two conceptions differ completely and are associated with opposite philosophies of the external world. The other great discovery of the present century is the quantum theory, which in its perfected form of quantum-mechanics appeared in 1925: this also is completely irreconcilable with the postulates of Newtonian science.

We have therefore come now to the end of an age—the age of classical physics—which we may count as having extended from the publication of Newton's *Principia* in 1687 to the acceptance of General Relativity and Quantum-Mechanics by our own contemporaries—about 250 years. The replacement of the Newtonian fundamental assumptions by a wholly different set of concepts, which is now taking place, represents a change in the philosophical view of the world which future generations will regard probably as one of the major turning-points in the history of thought, and perhaps as the most significant event of our time.

This afternoon I propose to set the revolution of the twentieth century, by which the doctrines of classical physics have been overthrown, side

* Address of the President at the Annual Statutory Meeting of the Royal Society of Edinburgh, October 26, 1942.

† Communicated by the Author.

by side with the revolution of the seventeenth century, by which they were originally established. It will appear that in some respects the second movement is reversing the direction of the first, and is bringing back ideas which were accepted in the great days of the Scholastic philosophy, but which, having been discarded at the Renaissance, have for the last three hundred years been unknown outside a small circle of scholars.

I have therefore taken for these remarks the title *Aristotle, Newton, Einstein*. "Aristotle" stands for the interpretation of the world which was developed, on the foundation of the Aristotelian physics and metaphysics, by the Scholastic philosophers of the thirteenth century; "Newton" symbolizes classical physics, which displaced Scholasticism and is now in its turn outmoded; and "Einstein" represents the new conceptions which have arisen in connexion with quantum-mechanics and general relativity, and on which the physics of the future must be based.

Let us, then, first inquire what was gained and lost when the mediæval philosophy was superseded in the seventeenth century by the doctrines of Descartes and Newton. To answer this question it is necessary to examine how the supersession came about. From the fourteenth century onwards Scholasticism was decadent, and by the end of the sixteenth it had become thoroughly debased. The love of Nature that had been so vital in Aristotle had almost perished; the practice of observation and experiment, on which he and St. Thomas had so strongly insisted, was neglected save by a few solitary workers; and the degenerate Schoolmen occupied themselves with futile subtleties that bore no relation to life and reality; they argued about homogeneities and heterogeneities, categorematics and syncategorematics, and showed the causes of things in sympathy, antipathy, and the influence of the heavens*. No wonder that the virile scholars of the Renaissance broke away from it all. In Italy, under the patronage of the Medici, there was a revival of Platonism; and at Paris, in 1536, a crowded audience acclaimed the thesis of Peter Ramus, "*Whatever is in Aristotle is false.*"

If philosophy and science were to be restored to life and health, the first necessity was, as Ramus saw, to re-establish contact with the external world. The chief pioneer in the movement back to Nature, the great exemplar of the accurate quantitative observation of phenomena, was a friend of Ramus's, the Danish astronomer Tycho Brahé, who lived from 1546 to 1601, and was thus about twenty years senior to Bacon and Galileo, and a century earlier than Newton. His observations, though made before the invention of the telescope and the micrometer, were

* It should in fairness be mentioned that a great part of the principles of dynamics had been discovered by scholastics in the University of Paris in the fourteenth century, and that their doctrines, though they did not appreciably influence the general body of scholastic teaching, were known to Leonardo da Vinci and Galileo, who built on them.

astonishingly accurate ; and some of his deductions from them were soon seen to be incompatible with the Aristotelian system of the world ; thus, his memoir on a new star which appeared in the constellation Cassiopeia in 1572, by showing that this body was situated among the fixed stars, destroyed the belief in the eternal incorruptibility of the heavenly bodies ; and his proof that the comet of 1577 moved round the sun in planetary space, shattered the cosmology which located comets in the earth's atmosphere.

The work of Tycho firmly established the principle that natural philosophy must be based on quantitative data acquired from observation. But something more was needed in order to consummate the foundation of modern science, and this further element was contributed by his pupil Kepler. Kepler derived his conception from the revived Platonism which was then in favour ; but actually it is traceable beyond Plato to his predecessors the Pythagoreans, and it may have been due to Pythagoras himself.

The original Pythagorean discovery related to the lengths of the strings of a lyre : it was found that if a string is stopped at half its length it gives a note one octave higher ; if at two-thirds its length, it gives a note higher by the musical interval called a fifth, and so on. Thus simple numerical ratios were shown to exist between the lengths of strings which produce sounds harmonious to each other, and so a connexion was set up between mathematics and æsthetics. This was generalized into the principle that numerical laws, analogous to the numerical laws of harmony in music, were the proper means of interpreting the fundamental unity of the cosmos ; that there must be a mathematical harmony of the external world underlying all phenomena ; that this was the reality which philosophers sought, and that the task of men of science was to find it.

Moreover, it was asserted that the dispositions of Nature were of the simplest character that could be imagined in any particular case. This consideration, which has in fact played a part of the first importance in the history of physics—in our own day it guided Einstein to the law of gravitation in curved space—was applied by Kepler in order to simplify the elaborate picture of the world which he had inherited from his predecessors. It is to be remembered that Copernicus, although he took the all-important step of placing the sun in the centre of the universe, still retained the intricate machinery of epicycles which had been devised by Hipparchus to represent the motions of the planets, and which, by the successive adjunction of fresh curves to represent new discoveries, had by now become intolerably complicated : so much so that a royal patron of science, to whom it was described, remarked that “ if the Deity had consulted him at the creation, he would have given Him good advice.”

It seemed to Kepler that the truth must be much simpler than anyone had yet realized ; and that by use of the right kind of mathematics it should be possible to exhibit or suggest in some way a physical connexion between the planets and the sun as the centre of their motions. Even-

tually he succeeded in showing that the planes of the orbits of all the planets pass through the sun ; that all the orbits are ellipses, having the sun as a focus ; and that a line joining a planet to the sun sweeps out equal areas in equal times.

By the labours of Tycho and Kepler the modern procedure of science was instituted, and the true structure of the universe was revealed. At this point it may be observed, that while the Scholastic cosmology * was thereby completely disproved and overthrown, there was nothing in the new methods and discoveries that was inherently irreconcilable with the Scholastic metaphysics : the whole of Tycho's and Kepler's work might conceivably have been absorbed into the philosophy of the Schoolmen by a peaceful and conservative revolution. If this had happened, we in the twentieth century should have been spared the necessity of readjusting our position by a movement back towards Aristotelianism ; but it was not to be. What did happen was a violent upheaval which swept away ontological doctrines equally with cosmological, destroying the old order entirely ; an upheaval out of which the system of classical physics was formed, and which has dominated the relations of science, philosophy, and religion with each other down to the present time.

The central figure of the movement on its metaphysical side was Descartes. As a young man he had become familiar with the degenerate Scholasticism of the day ; but it had left him dissatisfied. Its conclusions were based principally on the affirmations of the great doctors ; but the authority of the doctors was insecure, and the only branch of knowledge that seemed to be satisfactorily established was mathematics, whose procedure was to set out from self-evident postulates and to deduce from them results of practical value and incontrovertible truth. Descartes conceived the idea of searching for principles as certain as the axioms of mathematics, and on them as foundation to rebuild philosophy.

In pursuance of this design, he proposed far-reaching changes in the philosophy of Nature. The first step—evidently suggested by the success of Kepler's work on the planetary orbits—was to describe the happenings of the external world in mathematical language. Now of all things presented to our observation, the spatial dimensions of bodies are the most obviously quantitative ; he therefore seized on this feature, and based his system of the world on the affirmation that *the characteristic of matter is extension*. Another experience which is measurable is the passage of time ; and hence the movement of bodies, which may be specified by the distance passed over in intervals of time, also admits of quantitative treatment. In terms of these two concepts, matter and movement, Descartes proposed to explain the physical universe : quality was to be made intelligible as varying quantity.

In this scheme, extension constitutes matter, and matter constitutes space, which is therefore a plenum—there is no void. The sensations of

* The word *cosmology* is here used in the sense that is customary in modern scientific writing, not the wider sense in which it is used in Scholastic philosophy.

sound, light, heat, taste, and qualities generally, are to be regarded as belonging to our consciousness, and purely subjective : in Nature itself there is nothing but extension and the locomotion of its parts ; the external world is a purely mechanical assemblage.

In the Cartesian transformation of philosophy the very meanings of the keywords were altered. Thus *motion*, which to the Scholastics had meant change of any kind, was now restricted to mean change of position ; *matter*, which in the older doctrine was correlative to *form*, now meant simply corporeal being. Especially noteworthy is the new importance acquired by *Space* and *Time*. The Schoolmen had no word for "space" as we understand it ; for *spatium* had rather the sense that "space" has in the Authorized Version of the Bible, *e. g.* "All with one voice about the space of two hours cried out, 'Great is Diana of the Ephesians'"* ; while *locus* meant the space occupied by a particular body. *Where* and *when*, which to the Scholastics had been merely two among the ten predicaments of Being, now came to dominate completely the description of Nature.

That description was even more strictly mechanical than the Newtonian description which later superseded it, as may be seen, for instance, in their respective conceptions of gravitation. Gravity had been classified by the Schoolmen as an "occult quality"—that is to say, a force or tendency produced by no visible agency. Descartes denied the existence of occult qualities, and maintained, like the Greek atomists, that impact was the only mode in which one body could affect another : consequently he was compelled to furnish a new explanation of the fall of bodies towards the earth. This he did by postulating that surrounding the earth there is a vortex of subtle matter, or æther, which, by its pressure, provides the effect of gravity. Newton, on the other hand, formulated the inverse-square law without providing any mechanism to account for it ; and in the Preface to the second edition of the *Principia* there is a frank reversion to the Scholastic view of gravity as an occult quality.

Thus in the picture of the world arrived at by Descartes, all the phenomena of astronomy and physics, so far as they were known at the time, were represented by aggregations or motions or pressures in the plenum of space. Nothing resisted his mechanical explanations, except the thought of man ; this could not be brought into any relation with extension, and was evidently not amenable to mathematical analysis. It must, therefore, he concluded, have a principle other than matter ; and thus he arrived at a dualistic philosophy, and divided reality into the two great classes of extended and thinking substances, *res extensa* and *res cogitans*, the objective and the subjective, the corporeal and the spiritual world. As matter is characterized by extension, so the mind is characterized by thought : the two are completely independent, and no explanation of any relation between them is forthcoming.

* Acts xix. 34.

The complete disjunction of the psychical from the physical, which was characteristic of Cartesianism, has profoundly affected the subsequent history of science, and indeed of almost every department of human thought. In the first generation after Descartes there was an uneasy recognition of the possibility that—since any view of the cosmos must have a theological bearing—the new natural philosophy might prove harmful to religion: and in fact a keen controversy broke out on this very question. The dispute was centred round the doctrines of Space and Time, which in this period underwent a profound change.

The principal agent of the change was Pierre Gassendi (1592–1655), who, opposing Descartes' representation of space as a plenum, revived the doctrines of the ancient atomists regarding the void. This implied making a distinction between matter and extension, and asserting that while space has extension, matter has solidity as well, and occupies only a part of space. Gassendi's ideas were adopted by Newton, and thus was evolved the portrayal of space and time which became finally established in classical physics. Its fundamental postulate is that all the phenomena of the external world can be described in terms of the location and motion in space of entities, each of which has some degree of persistence and continuous identity in time. Thus whatever happens, happens in space—space, the stage on which the drama of physics is to be played, is the dominating conception of the whole system. This had not been at all the point of view of Scholasticism, and, as we shall see later, the recent progress of physical discovery has shown that it is radically unsound.

By Newton, space is regarded as having a positive, objective existence, which is not attached in any way to subjective necessities of the human mind. "Absolute space," he says, "in its own nature, without regard to anything external, remains always similar and immovable" * and "all things are placed in space as regards order of situation." † It is a real entity, subsisting prior to, and independently of, the bodies which it contains; and all events in Nature can be represented by movements within it. Every point of space persists throughout an infinite succession of instants of time, and the notion of *simultaneity* is valid, with all the implications which it carries in classical physics. In the Gassendi-Newton scheme it is not considered necessary to account for the existence of entities which are permanent over appreciable durations of time, such as (in the earlier period) particles of matter: these are postulated, and the aim of science is to explain the changing phenomena of the universe in terms of their motions. Persistence of bodies in time, and their displacement in space, are the concepts to which everything in the external world is to be reduced.

The attempt to fit this doctrine into the framework of philosophy and

* "Spatium absolutum, naturâ suâ sine relatione ad externum quodvis, semper manet simile et immobile," *Principia*, Schol. ad. Defn.

† "In spatio quoad ordinem situs locantur universa," *ibid.*

theology was confronted by the difficulty that besets all systems based on the Cartesian bifurcation between mind and matter, namely, that no provision is made for the interaction of spiritual with corporeal being. A possible solution seemed to be indicated by an idea put forward in 1647 by the Cambridge Platonist Henry More, that in some part of the human brain there is a *sensorium* or organ of internal sensation where the understanding resides, to which the images of external things are conveyed by the organs of sense, and where they have a "tactual conjunction" with the soul, which thus perceives them. Newton now boldly suggested that Space might be the Sensorium of God. "Does it not appear from phenomena," he said*, "that there is a Being incorporeal, living, intelligent, omnipresent, who in infinite space, as it were in his sensory, sees the things themselves intimately, and thoroughly perceives them, and comprehends them wholly by their immediate presence to himself?"

The idea that God had to be fitted into a scheme of which matter, space, and time were the primary concepts, was attacked by Leibnitz who rejected altogether the doctrine of an absolute space and time having reality outside our minds, and maintained that space is only a conceptual entity, an order according to which situations are disposed, and time is only an order of succession. His argument against Newton may be put in a modern form somewhat as follows: During the operation of "summer time" the clock is an hour ahead of Greenwich time. This fact is, however, not made evident by any of the ordinary happenings of life, since all clocks, departures of railway trains, office hours, mealtimes and so forth, bear to each other the same relations as before: in order to detect the change we have to observe something which does not obey the Act of Parliament establishing summer time, such as the moment of sunset. Now suppose that some way could be found of compelling the heavenly bodies to adapt themselves to summer time on the same day as our clocks: then, after this, it would be impossible by any observations whatever to tell which kind of time we were keeping: the only evidence would be that furnished by memory—the recollection of the day which had only 23 hours, when the clocks were put forward. Let us now imagine that day to recede into the past, back to the creation of the world. Would there then be any difference between the two systems? Or to put the same question in another form, is there any meaning in the statement that God might have created everything an hour sooner? Newton would say, Yes. Leibnitz would say, No. I leave the Fellows of the Society to form their own judgment on the matter.

Another count in Leibnitz's indictment of Newtonianism related to the concept of *force*, the *vis motrix* of the *Principia*, which in the case of gravity was represented as acting at a distance. "Some men," wrote Leibnitz, "begin to revive, under the specious name of *forces*, the *occult*

* *Opticks*, Qu. 28.

qualities of Scholasticism; but they bring us back again into the Kingdom of Darkness." * Now "force" in its statical sense, as, for instance, when we speak of "the force exerted by the weight of one pound," was a familiar idea to the Schoolmen: their physics included what they called *scientia de ponderibus*, which dealt with such matters as the law of equilibrium of the lever and the apparent weight of a body resting on an inclined plane. But the kinetic relations of force were unknown in the Middle Ages, and were first formulated in the *Principia*, in the Second Law of Motion, which equates force to the product of mass and acceleration.

Newton showed how the known connexion between forces and accelerations made it possible to write down equations, by which the motion could be determined. But before long it was found that, at any rate when the system studied was only a single particle, the motion could be calculated without bringing in forces or accelerations, by making use of what would now be called the principle of conservation of energy. When a pendulum is drawn to one side, so that the bob is higher than when it is hanging vertically, it has energy in the potential form (in this case due to gravitation); when the pendulum is released, it begins to swing, the potential energy being changed into energy of motion. This way of regarding dynamics shows a certain affinity with the Scholastic philosophy; for the physics which the Schoolmen appropriated from Aristotle was dominated by the idea that there are two ways of Being, potency † and act ‡, and that all change is a transition between a state which is potential and a state which is actual.

For a problem in classical mechanics there is of course no contradiction between the solution which is obtained by using the concept of potential energy and the solution by the original Newtonian method which uses the concept of force: the two are mathematically equivalent. But in the present century it has been found that the mutual influence between elementary particles is governed not by the laws of classical mechanics but by those of quantum mechanics, in which the concept of force is abolished and the interaction is represented by a term corresponding to the potential energy of classical mechanics, but of a more abstract character; it is called an *operator*, and, like potential energy, may be attached to the Aristotelian concept of potency.

To return to classical physics, a further change in its outlook was brought about by the discovery of what are called *minimum-principles*. A simple example is afforded by the hanging chain. Suppose that a chain is suspended between two points of support: what will be the curve in which it will dispose itself? We all know the general appearance of the curve: it is something like an arc of a circle joining the two points of suspension. Actually it is not part of a circle, but a curve known as

* A Collection of Papers which passed between Mr. Leibnitz and Dr. Clarke. (London, 1717), p. 265.

† δύναμις, *potentia*.

‡ ἐνέργεια or ἐντελέχεια, *actus*.

a *catenary*: this was discovered towards the end of the seventeenth century by the Newtonian method of considering the forces which act between consecutive links of the chain. The problem may, however, be solved in a quite different way, without introducing the idea of force, by assuming that the chain arranges itself in such a way as to make its centre of gravity have the lowest possible position. This assumption, that the height of the centre of gravity is a minimum, is a typical minimum-principle. At least one minimum-principle was known to the ancient Greeks, namely, that when a ray of light issuing from a source is reflected at a mirror and afterwards received by an observer, then the path followed by the incident and reflected rays is shorter than any other path from the source to the observer which meets the mirror.

By a succession of discoveries it was shown that all the happenings of Nature can be predicted by means of minimum-principles: the climax was reached in 1915, when Hilbert showed that all physical events (gravitational, electrical, etc.) in the universe are determined by a "world-function" which is such that its integral taken over the whole of space-time is a minimum. Such a statement as this has a decidedly Aristotelian character; for, in studying change, Aristotle always fixed his attention on the end to be fulfilled: his science was essentially teleological. Again we see how classical physics, following its own natural development, tended to deviate from the pattern devised by Newton, and to return to the Aristotelian-Scholastic mould.

A change in orientation such as this, however, did not affect the fundamental assumptions of Newtonian science. We have now to consider developments of a far more subversive nature.

The first serious trouble arose in connexion with the doctrine of space. The space of Gassendi and Newton was, so far as geometry was concerned, the space of Euclid: it was infinite, homogeneous, and completely featureless, one point being just like another; so far as physics was concerned, it was like the vacuum of the ancient atomists, mere emptiness into which things could be put. From the philosophical point of view this concept was open to the objection that Aristotle had urged against the doctrine of the atomists, namely, that if space were devoid of local properties, the tendency of a body to move spontaneously in a particular direction (*e.g.* the existence of a gravitational field, as we should say) would be unintelligible. As a matter of fact, the successors of Newton felt this difficulty; and, having started with a space which was in itself simply nonentity, having no property except a capacity for being occupied, they proceeded to fill it several times over with æthers designed to provide electric, magnetic, and gravitational forces, and to account for the propagation of light; and as it was impossible to draw any effective distinction between these æthers and space, Newtonian space became eventually a plenum of the most elaborate kind, possessing such qualities as density and rigidity everywhere. Its points were capable of individual identification, and could be regarded as fixed; and having thus acquired a more definite and concrete substantiality than

Newton himself had ever contemplated, its absolute character became an essential and inseparable axiom of classical physics. But the discovery in 1905 of the principle of relativity led to inferences incompatible with the existence of any kind of quasi-material æther : and thus the Gassendi-Newton doctrine became involved in hopeless contradiction.

The problem that now confronted physicists was this : how can local properties, such as a gravitational field, exist in space, when the existence of an æther is not a permissible supposition ? The answer was furnished, in 1915, by the *General Theory of Relativity* of Einstein. He discarded Gassendi's assumption that space was a uniform characterless vacuum, and postulated that it had a property of *curvature*, varying from point to point : and that just as (to make use of a rough analogy) a paramagnetic body when placed in a magnetic field tends to move from the weaker to the stronger places in the field, so a massive body in space might be pictured as moving from places of weak to places of strong curvature. The curvature, in fact, performs in General Relativity the same kind of function as the density and rigidity of the æther did in classical physics ; but, unlike the æther-properties, it does not come into conflict with the principle of relativity. In Einstein's conception, space is no longer the stage on which the drama of physics is performed : it is itself one of the performers ; for gravitation, which is a physical property, is entirely controlled by curvature, which is a geometrical property of space.

In Einstein's theory of gravitation the Newtonian concept of force is completely done away with ; a free particle moves in a path determined solely by the curvature-properties of space ; it is, as the Aristotelians would say, *in potency* with regard to space, and things in a state of potency continually seek to become actualized. The changes of position of the particle, in their turn, bring about changes in the curvature of space, so that the particle and space together may be regarded as a single system whose evolution is determined by the law that the total curvature of space-time is to be a minimum : as we may say, *gravitation represents a continual effort of the universe to straighten itself out*—a statement so completely teleological that it would certainly have delighted the hearts of the Schoolmen.

While classical physics was thus being undermined by the principle of relativity, an even more devastating attack on it was developing from the side of quantum-theory. In 1913 a young Danish research student named Niels Bohr, working under Rutherford in Manchester, published some new and revolutionary ideas regarding the way in which light is generated. Let us take as the simplest example the generation of light by a hydrogen atom. This atom consists of a massive particle in the centre, with an electron circulating round it, just as the earth and the other planets move round the sun. Now in the solar system a planet may revolve at any distance whatever from the sun—that is to say, there is no restriction on the dimensions of the planetary orbits ; and similarly, according to classical physics, the electron in the hydrogen atom might

revolve at any distance whatever from the nucleus—the possible orbits would form a continuous sequence. Bohr, however, now put forward the suggestion that this deduction from classical physics is false, and that, in fact, only certain particular orbits are allowable ; just as if in the solar system a planet could move in the orbit of the earth, or in that of Mars or Jupiter, but could not move in any orbit intermediate between these, such as the orbits in which the minor planets actually do move. When the electron is moving in one of the permitted orbits the atom is said to be in a *stationary state* ; and the fundamental assumption of Bohr's theory is that the atom, when it is not emitting light, must always be in one or other of the stationary states, without the possibility of its being in any intermediate condition : the emission of light takes place when, and only when, the atom changes from one stationary state to another.

Bohr showed that his suggestion would explain many of the known features of spectra most admirably ; but some serious objections could be brought against it, and one of them was that he could give no explanation of the process by which an atom in a state A is raised or lowered to another state B. In the change the electron must transfer itself from the orbit belonging to state A to the orbit belonging to state B ; and according to the Gassendi-Newton doctrine, an object such as an electron can transfer itself from one position to another position only by travelling over the space between them, occupying in succession the whole continuous sequence of intermediate positions. Bohr, however, found it impossible to provide any description of the transference of the electron, and was compelled to renounce the attempt to explain transitions between stationary states. At the time, this was regarded as an imperfection in his work. In the light of our knowledge to-day, we take a very different view, and regard Bohr's renunciation as one of the most valuable and permanent features in his theory, and as a landmark in the history of science ; for the subsequent development of quantum-mechanics has shown that his inability to trace the adventures of the electron between leaving orbit A and settling in orbit B was not due to any insufficiency on his part, but was inherent in the physical situation.

What is the difficulty ? Is it that the mathematical operations required to calculate the motion are so intricate as to be beyond the skill of the best mathematicians ? Or is it that owing to the imperfections of our laboratory apparatus we cannot make measurements of sufficient delicacy to specify empirically the successive stages of the motion ? No, the trouble is more deeply seated. Even if we could imagine an investigator capable of solving any set of mathematical equations whatever, and, moreover, possessing instrumental equipment of the highest refinement conceivable, even then the problem of depicting in terms of the concepts of classical physics the transition of the atom from one stationary state to another would be insoluble ; and the reason is, that the process cannot be described as a continuous movement of an electron in space. We are confronted by a theoretical impossibility, like the

impossibility of expressing π as a rational fraction, or the impossibility of constructing a regular heptagon with ruler and compasses.

The importance of this discovery is that it invalidates the pre-suppositions of the whole Gassendi-Newton doctrine. It shows that there are events in the physical world which cannot be represented on the background of space and time.

It therefore becomes necessary to find a metaphysics different from that which has been associated with classical physics; for metaphysics must (as Aristotle held) originate with reference to physics, since it is the conceptual framework into which our experience of Nature is to be fitted. The progress of science has destroyed the foundations on which natural philosophy has hitherto been grounded. How is the damage to be repaired?

Evidently space and time must be deposed from the dominant position which they held in Newtonianism, and relegated to a status more or less resembling that which they had in the Scholastic philosophy; and therefore we must now begin not with space and time, but with fundamental physical events, such as the one which has just been occupying our attention, namely, the assumption by the atom of its different possible stationary states. The atom, which has a potency of various states, is correlated to the states, as potency is to act. It endures as the atom, while it takes different states in succession. This is precisely the aspect of things on which Aristotle fixed his attention: that substratum which persists while receiving different determinations is what in the Aristotelian-Scholastic philosophy is called *matter*; whereas the structural principle, which is peculiar to each determination or state, is called *form*. The atom, then, is *matter* with respect to its states, which are *forms*.

The same principles can be applied in one of the most recent of physical theories, that of the nucleus of the atom. The nucleus is generally said to be composed of two kinds of elementary particles, called *protons* and *neutrons*; but in the quantum-mechanical theory of processes such as the emission of beta-rays, the proton and neutron are regarded as two "states" of a single entity, often called a "heavy particle." Here, then, the heavy particle would be "matter," and its determinations as a neutron or proton would be its two possible "forms."

Matter is correlated to form, as potency to act—notions which may again play an important part in the natural philosophy of the future, as they did in pre-Newtonian days.

In the light of the Aristotelian-Scholastic concepts, certain otherwise puzzling facts, which have been discovered in modern investigations, fall into their places as elements of a rational coherent system. Take, for example, the fact that all electrons have the same electric charge. For this, the classical theory of electricity has no explanation to offer: the law of interaction is that two electrified particles repel each other with a force proportional to the product of their charges and inversely proportional to the square of the distance between them; and this law

is valid whether the charges are equal or unequal. Yet it is impossible to believe that the actual equality of the charges of electrons is a mere accident : it must be fundamental in the scheme of Nature, and there must be a reason for it. How fundamental and necessary it is, has been shown by a study of the forces by which atoms are held together so as to form molecules. Take, for instance, the hydrogen molecule, which is constituted of two hydrogen atoms, each atom consisting of a nucleus and an electron. If the two electrons are interchanged with each other, there is no change in the system, since the electrons are identical. From which feature of the situation, by following up its consequences in the light of quantum-mechanics, we can predict that a stationary state of the system exists which has less energy than the energy of the two atoms when separated : this stationary state corresponds to the stable hydrogen molecule.

In the proof, everything turns on the exact equality of the two electrons. If they had different charges, the binding force (which is purely quantum-mechanical) would not exist ; and this is true of all those binding forces which are called " homopolar bonds " in chemistry. Thus the world would be a very different place from what it is, if all electrons were not identical.

But the matter is not yet exhausted. There is something more profound : electrons are indistinguishable in a still more rigorous sense. If two electrons are at one instant at places A and B, and at a later instant at places C and D, it is impossible to say which of the electrons at C and D is the one which was formerly at A—that is to say, an electron can freely exchange its recognizability with other electrons ; it has no sameness of being, no proper identity, no separate history. Its selfhood is merged in an electronhood which it shares with all other electrons, and which is correlated to it as potency to act. From the philosophical point of view this is clearly important, for it necessitates a revision of the concept of individuality as applied to the elementary particles, and reopens, in connexion with the most recent discoveries in physics, the question which engaged so much attention in the Middle Ages, regarding the nature of *universals* or general terms, which represent the common basis of a class of individual objects.

The transition which is now in progress from classical physics to a new natural philosophy conformable to relativity and quantum-mechanics is less violent than that which took place three centuries ago, when classical physics arose on the ruins of Aristotelianism, but it may prove not less significant. As we have seen, it involves a return to some fundamental Aristotelian notions, which have again become a living force ; and this should lead to more intercourse and mutual understanding between men of science and philosophers ; for of all types of philosophy, the Aristotelian-Scholastic is, in its principles, the most congenial to the scientific mind. Like men of science in all ages, the Schoolmen never doubted the existence of an objective world that was independent of human cognition ; they were untroubled by difficulties such as those

raised later by Berkeley, Hume, and Kant; they looked outwards towards a reality external to themselves, and analyzed their experience of it. They held that *all* our knowledge is derived primarily through the senses, which come into direct contact with concrete things. The sense-impression is subjected to the operation of the active intellect, which throws light upon it, as it were, divesting it of its contingent elements and making it intelligible by drawing out the idea or concept contained in it. The idea so obtained is the means whereby knowledge is acquired. Thus the human intellect is capable of conceiving relations such as cause and effect, and of apprehending Being as such; metaphysics is possible, and completes physics by ascending to the true understanding of reality.

All this fits in very well with the scientific man's view of what metaphysics ought to be. But if the prospect of a movement in the direction of Aristotelianism is agreeable to the investigator of Nature, it may prove not less so to the philosopher. For the Cartesian revolution, which dethroned Aristotle, severed the philosophic and scientific traditions from each other, and made it impossible to incorporate physics into an all-embracing doctrine of Reality. The impoverished representation of the objective world which Descartes obtained by abstracting only its purely quantitative aspects was a soulless mechanism, composed of parts which had no function except to move each other about in space; and this function was itself philosophically inexplicable, and had no relation to any ideas of value or purpose.

The inherent defects of Newtonianism, the result of its dependence on the concepts of Descartes and Gassendi, were perceived by Leibnitz. In his controversy with Clarke he discussed the tendency, which had become common in Newtonian circles, to conceive of the relation between God and the universe as analogous to that of a watchmaker to a watch which he has constructed, and which, having been set going, continues to function, for some time at any rate, without any necessity for the continued presence or attention of its originator. Such a conception led inevitably to the idea of an absentee God, who, having created the world, had left it to run its own course without further divine intervention, and who was therefore for practical purposes non-existent. As Leibnitz saw, it is impossible to build any religion as a superstructure on a purely mechanical philosophy; and, in particular, Christianity, being an incarnational and sacramental religion, is incompatible with any view of the world which completely despiritualizes matter.

The debate between Leibnitz and Clarke took place in the lifetime of Newton, who, however, did not participate in it. Though profoundly interested in theology, he seems to have held that the physicist is not under any obligation to concern himself with metaphysics; he can give his undivided attention to investigating the laws which will enable him to predict phenomena, and can leave the deeper problems entirely out of account; he can make it his purpose to describe rather than to explain. This was one of the implications of his celebrated declaration *hypotheses*

non fingo *, and it determined the attitude of his successors—that is to say, men of science since Newton have generally held that correct (even if in some respects limited) knowledge regarding physics can be combined with any views whatever on the fundamental questions of Being and Reality; that part of the world can be rightly understood without reference to the whole; that natural philosophy is independent of metaphysics.

In a restricted sense this doctrine is true. The fact cannot be disputed that great discoveries regarding the behaviour of the external world have been made by workers whose investigations in their field of research were not related in their own minds to any interest or belief outside it. But the effect of such segregated thinking has been to make science a departmental affair, having no influence on life and thought except indirectly through its applications. At the present time there is a movement in scientific circles aiming at securing for science a greater influence on human affairs, and even calling for a refounding of civilization on a scientific basis; but its advocates do not always understand that, as a necessary condition for the possibility of such a reform, science must be reintegrated into a unity with philosophy and religion.

XXXI. *A Note on the Analysis of the Nuclear Binding Energies.*

By VLADIMIR VAND, D.Sc. †

[Received November 4, 1942.]

ABSTRACT.

The function $L(A)$, appearing in Wigner's theory has been calculated by two different methods: from the observed mass defects and from the chemically determined atomic weights of the elements. It has been found that most probably the curve $L(A)$ is not a smooth function of A , but the accuracy of the individual values of L is insufficient to give the exact shape and character of the irregularities. An empirical two-constant formula, disregarding the oscillations, is suggested. The discrepancy of the values in the region where the two methods overlap is probably caused by the fact that the most abundant nuclei have higher proton content than the most stable ones, which supports the view that the nuclei were formed from the proton side of the stability curve.

THEORETICAL formulæ for the nuclear binding energy, as a consequence of the application of the Pauli principle to a many-body simplified nuclear

* *Principia*, Schol. gener. sub finem.

† Communicated by the Author.

model, have been given by E. Wigner ⁽¹⁾ and were successfully used by W. H. Barkas ⁽²⁾ for computation of about 150 masses of both known and as yet unknown nuclei. The formula, as used by Barkas is

$$M = 1.00853 A + 0.0008 T_{\xi} + E_0(A) + \mathcal{E}' L(A) - 0.000638 A^{-1} (A - T_{\xi} - 1) T_{\xi} + 0.0264 T_{\xi}^2 / A, \quad (1)$$

where M is the atomic mass measured in terms of $0^{16} = 16$, A is the mass number, $T_{\xi} = \frac{1}{2}(N - Z)$ is the isotopic spin of the nucleus, N is number of neutrons, Z is number of protons forming the nucleus, $E_0(A)$ and $L(A)$ are functions of A , appearing in Wigner's theory, and the factor $\mathcal{E}'(A, T_{\xi})$ is a number which represents the ground state symmetry character of the configuration and is a periodic function of A with a period of 4. Wigner gives general rules for calculation of the values of \mathcal{E}' and Barkas gives a table of values of \mathcal{E}' for light nuclei.

For our further purposes it will be more convenient to replace the factor \mathcal{E}' by a somewhat simpler factor Ψ , according to the equation :

$$\mathcal{E}'(A, T_{\xi}) = \frac{1}{2} + \frac{1}{2}(|T_{\xi}| + 2)^2 + \Psi(A, T_{\xi}). \quad (2)$$

The numerical values of the factor Ψ are given in Table I. (k is a whole number.)

TABLE I.
Values of the factor Ψ .

A	$T_{\xi} = \pm 0$	$\pm 1/2$	± 1	$\pm 3/2$	± 2	$\pm 5/2$	± 3	$\pm 7/2$	± 4
$4k$..	0	—	$3/2$	—	0	—	$3/2$	—	0
$4k+1$..	—	$3/4$	—	$3/4$	—	$3/4$	—	$3/4$	—
$4k+2$..	$5/2$	—	0	—	4	—	0	—	$3/2$
$4k+3$..	—	$3/4$	—	$3/4$	—	$3/4$	—	$3/4$	—

From equation (2) and from Table I. it is immediately apparent that Ψ is again a periodic function of A with a period of 4 and that in addition the factor Ψ has the following simple properties :—

- (1) For the most abundant nuclei of the Z even, N even type it is $\Psi = 0$.
- (2) For the medium abundant nuclei of the Z even, N odd or Z odd, N even type it is $\Psi = 3/4$.
- (3) For the scarce nuclei of the Z odd, N odd type it is $\Psi \geq 3/2$.

The meaning of the separate terms of equation (1) is as follows :—

The first term of equation (1) is the average of the mass of the neutron (1.00893) and of the hydrogen atom (1.00813) multiplied by the total number of particles. The second term allows for the mass difference between the neutron and hydrogen. More recent measurements of the mass of the neutron and of the H^1 affect slightly the value of the coefficients in the first and in the second term, which is, however, for our

purposes of no serious importance. The third term $E_0(A)$ represents the general trend of the binding energy of the nucleus as a function of A , and includes a linear term, due to the saturation properties of the nuclear forces, a surface energy $A^{2/3}$ term and the part of the Coulomb energy term, which depends on A only. The fourth term gives the dependence of the energy on the symmetry character of the nuclear configuration. The fifth term gives the Coulomb energy of the protons under the assumption of the $A^{1/2}$ law for the nuclear radius (the exchange part of the Coulomb energy and some other refinements, as discussed by Present ⁽³⁾ and others, were omitted for reasons of simplicity). The numerical factor of 0.000638 M.U.* of this term given by Barkas is in satisfactory agreement with the newer measurements published by Stephens ⁽⁴⁾ in the region of the heavier nuclei. The sixth term represents (at least approximately, as the kinetic energy especially of the lightest nuclei is not known with desired accuracy and cannot be represented by such a simple expression) the dependence of the kinetic energy on T_ξ and A .

Formula (1) can be used for the calculation of the functions E_0 and L from the observed mass defects. The functions E_0 and L being once known, the formula (1) can be used for the calculating of masses of unknown nuclei. Barkas found that the values of L do not check well below $A=12$ and that almost the entire useful range of the curve $L(A)$ can be represented by the hyperbola $AL=50$ m.M.U., in agreement with theory. The range of A investigated by Barkas ends with $A=50$.

As the function $L(A)$ may be found useful in the study of the other aspects of the nuclear theory, the computation of the function of $L(A)$ for the whole range of A was undertaken, using for lighter elements mass defects and equation (1), and for heavier elements a stability condition and abundance of isotopes.

For the light nuclei (up to $A=40$) the result obtained by a method of calculation, described by Barkas, is shown in fig. 1. A clearer picture of the behaviour of L is obtained if we plot the product $AL(A)$ against A , since the expected behaviour $AL=\text{const.}$ will show a horizontal line.

For the heavier nuclei above $A=30$ the function $L(A)$ can be computed from the abundance of the nuclei under the assumption that the most abundant nuclei correspond to the minimum of the atomic mass determined by differentiation of (1) with regard to T_ξ and by keeping A constant. We obtain thus

$$\begin{aligned} dM/dT_\xi = 0.0008 + L(A) dE'/dT_\xi - 0.000638A^{-1/2}(A - 2T_\xi - 1) \\ + 0.0528T_\xi/A = 0, \quad \dots \dots \dots (3) \end{aligned}$$

and since we can disregard scarce nuclei of the Z odd, N odd type in our computation, we can take Ψ in equation (2) as constant independent of T_ξ , and we can write

$$dE'/dT_\xi = T_\xi + 2. \quad \dots \dots \dots (4)$$

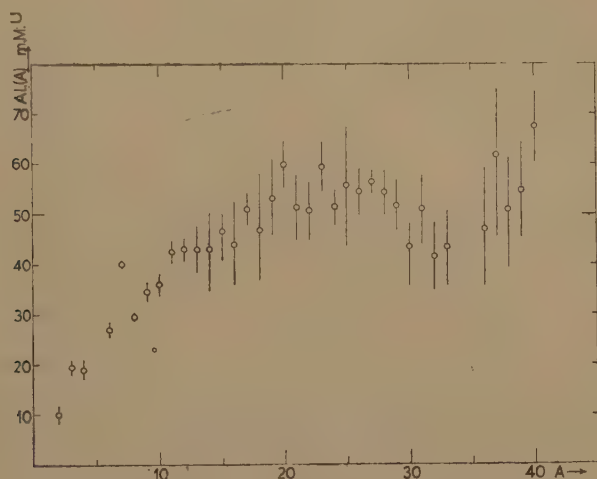
* 1 M.U.=1 mass unit=1000 m.M.U.=931.2 MEV.

By introducing this value in equation (3), we obtain the following equation for $L(A)$:

$$L(A)(T_{\xi}+2)=0.000638A^{-\frac{1}{2}}(A-2T_{\xi}-1)-0.0528T_{\xi}/A-0.0008. \quad (5)$$

We saw that equation (5) was derived under the condition of A constant. Its correct evaluation will be thus possible only if the absolute abundance of nuclei is well known. As the relative abundance of isotopes under the condition Z constant is far better known than the absolute abundance of elements, we can attempt to use the values of the maximum of the relative abundance of the isotopes instead of the maxima of the absolute abundance taken for A constant for the calculation of the function $L(A)$ from equation (5). The error so introduced will be smaller than the errors from other sources.

Fig. 1.



As the chemically determined atomic weights (after a slight correction for the mass defect and for the difference between the physical and chemical scale) actually represent weighted means already averaged according to the relative abundance of the isotopes as found in nature, we can use these for the calculation of the weighted mean of $T_{\xi} = (A - 2Z)/2$ and use the values of T_{ξ} so found for the calculation of $L(A)$ by means of equation (5).

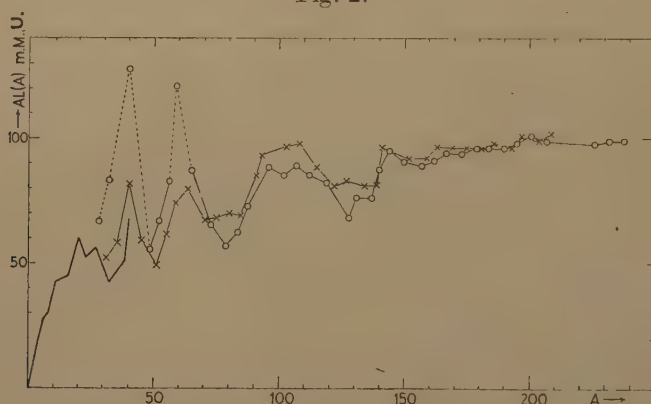
Results are shown in fig. 2, where the function $AL(A)$ is plotted against A . The values computed from atomic weights with even Z are represented by circles ; of elements with odd Z are represented by crosses. The values of fig. 1 are represented by a thick line for comparison. The accuracy of the method can be thus judged by the amount of discrepancy found between the values of fig. 1 and fig. 2, and by the amount of discrepancy between the curves connecting points of odd Z and even Z values.

Conclusions.

(a) Values of L computed by the two different methods show a relatively good agreement within limits of the possible error due to the crudeness of the second method in the neighbourhood of $A=40$. The curve for higher values of A makes a reasonable extension of the curve for small values of A calculated by the first method.

(b) The curve of $AL(A)$ shows departures from smoothness well above the probable error, with minima corresponding to the values of $A=16, 24, 32, 48, 80, 128$, which are probably connected with the filling of the various nuclear shells and changes of the radial extension of the wave function of the nucleus. The small accuracy of the individual values of L does not, however, allow a more detailed examination of the exact shape and character of the irregularities of the curve $L(A)$.

Fig. 2.



(c) The curve $AL(A)$ tends to an asymptotical constant value for very heavy nuclei in accordance with the theory. The value found is $AL(A)_\infty = 100$ m.M.U., which is about twice the value given by Barkas. The accuracy of the numerical results of Barkas, however, is not impaired, since in the interval of A investigated between 12 and 50 his value of $AL(A)=50$ m.M.U. lies just within the limits of experimental error.

As there is a possibility that the assumptions, under which formula (1) is derived, are not valid for the very heavy nuclei, we must accept the absolute values of L for heavy elements with a certain amount of caution. The constancy of $AL(A)$ for heavy elements is then rather surprising and perhaps significant.

(d) Sometimes for problems of a more general character it is useful to have some simple empirical formula for $L(A)$, which disregards oscillations of the function. A two-constant formula of the form

$$L(A) = \frac{100}{16+A} \text{ m.M.U.} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

can amply satisfy such requirements.

(e) Another very interesting point can be drawn from fig. 2. The discrepancy between the two curves calculated by the two different methods in the region of A, where they overlap, seems to be caused chiefly by the fact that our supposition, that the most abundant nuclei correspond to the minimum of the atomic mass does not hold, at least in that region. A systematic shift of the maximum abundance of nuclei towards the excess of protons seems to exist, *i. e.*, the most abundant nuclei have higher proton content than the most stable ones.

Such a deviation, if real, can be explained from the point of view of the evolution of the elements only.

It means, that at least in the region of the atomic numbers from $A=20$ to $A=40$ the formation of the nuclei proceeded from the proton side of the stability curve, probably by a successive emission of positrons, and few or none of the nuclei were formed from the neutron side of the stability curve.

The question seems to be of high importance, as it can throw light on the most fundamental process of element formation, supporting the view of the synthesis of the elements from protons (under suitable conditions of pressure and temperature) rather than their formation from neutrons or by a fission process from heavy nuclei.

References.

- (1) E. Wigner, Phys. Rev. li. p. 947 (1937).
- (2) W. H. Barkas, Phys. Rev. lv. p. 691 (1939).
- (3) R. D. Present, Phys. Rev. lx. p. 28 (1941).
- (4) W. E. Stephens, Phys. Rev. lvii. p. 938 (1940).

32 Temple Fortune Hill,
London, N.W. 11.
November 2, 1942.

XXXII. The Separation of Electricity in Clouds.

By Sir GEORGE C. SIMPSON, F.R.S.,
Acting Superintendent, Kew Observatory*.

[Received March 11, 1943.]

THE article on "The Separation of Electricity in Clouds" by Dr. J. Alan Chalmers, which appeared in the Phil. Mag. for January, 1943, so misrepresents my views on this problem that I must make an early reply. I regret having to do so at this time because I am at present engaged on the measurement and study of the electricity of rain, especially of non-thunderstorm rain, and would have preferred to wait until this study is completed before writing more on the subject.

* Communicated by the Author.

Chalmer's first sentence reads: "Simpson and Scrase⁽¹⁾ have put forward the theory that the main process of separation of charge within clouds is one of friction between ice particles of different sizes, the larger particles acquiring negative and the smaller positive charges." This is a complete mis-statement of the theory, but one so commonly made that I was at particular pains in my Presidential Address to the Royal Meteorological Society last year⁽²⁾ to set out in some detail the correct theory. The physical basis of the theory is the experiments made by W. A. D. Rudge⁽³⁾ and P. E. Shaw⁽⁴⁾, who found that when non-conducting materials are brought violently into contact and then separated they become negatively charged and the complementary positive charge goes into the air. This is true even when both colliding bodies are of exactly the same chemical composition and in the same physical state. The process was described in my Presidential Address in the following words: "He (P. E. Shaw) found that when identical bodies (insulators) are struck together they both become charged. Sometimes the two pieces obtain charges of opposite signs, but more often they both obtain charges of the same sign; almost invariably, however, the net charge obtained by the two bodies together is negative and prolonged blows in general increase the net negative charge. This, as he says, is most readily interpreted by supposing that positive ions escape to the air when the impacting solids separate." It will be seen that this process is not what is usually understood by friction electricity, in which one material becomes positively and the other negatively charged. The size of the particles has nothing to do with the process. I may mention in passing that the justification for assuming that this process takes place in thunderstorms is that it accounts for two other phenomena of atmospheric electricity, namely, the high electrical fields met with in tropical dust storms and in Antarctic blizzards.

Chalmers writes: "Simpson and Scrase⁽¹⁾ have, apparently, assumed that there are no water drops at temperatures below the freezing-point, but only ice particles." Considering what I have published on supercooled water drops in the atmosphere⁽⁵⁾ and on the icing of aircraft⁽⁶⁾, this is a truly amazing statement and I have searched through our paper to find on what it could possibly be based. I can find only the following: "Our observations have shown quite conclusively that the boundary between the positive electricity in the upper part of the cloud and the negative electricity in the lower is in every case in a region of the cloud where the temperature is well below the freezing-point and generally below -10°C . In this part of the cloud raindrops cannot exist. The cloud particles may be supercooled water, but on coalescing they would immediately freeze. The precipitation in the upper part of a cloud is in the form of crystals, either needles or plates. . . ." This does not say that there are "no water drops" but only that there are no "rain drops," and definitely says that supercooled cloud particles may be present. Nor do we say that there are "only ice particles," what we say is that "the precipitation is in the form of ice crystals." In such

a connexion precipitation has a technical meaning which excludes cloud particles.

In a later paragraph, Chalmers states "the observations by the alti-electrograph of Simpson and Scrase⁽¹⁾ and Simpson and Robinson⁽⁷⁾ have shown that the negative charge in a thunder-cloud is centred around the 0° C. isothermal, while the upper positive charge occurs at temperatures around -20° C." What are the facts? The question of the temperature at which the positive and negative charges occur is summarized in fig. 12 of the paper by Simpson and Robinson. There we find that the negative charges (with two explicable exceptions) all occurred at temperatures below 0° C., actually between 0° C. and -19° C.; while the positive charges occurred between -14° C. and -38° C. Whatever we may think about describing temperature between 0° C. and -19° C. as being "centred round the 0° C. isothermal" or temperature between -14° C. and -38° C. as "around -20° C.," we cannot approve of Chalmers taking the mean of these "round about" figures and saying "the results of Simpson and Scrase and Simpson and Robinson suggest that the separation of charges occurs at levels round about -10° C.," and then using -10° C. as though it were a definite temperature, in support of his own theory. As a matter of fact our observations show one case in which the separation occurred between -1° C. and -12° C. and another in which it occurred between -19° C. and -28° C., *i. e.* the two regions of separation had no temperature in common; hence temperature cannot be a determining factor. The only conclusion we were able to draw from our temperature observations was that, whatever the process may be which separates the positive from the negative electricity, it must be able to function where the temperature is well below the freezing-point; beyond this the observations give no indication that any particular temperature has any special significance. In the Antarctic we had blizzards with high electrical effects at temperatures below -30° C.

For the reason already given I do not wish at present to discuss Chalmer's suggestion that electrification takes place when water freezes—a suggestion which is by no means new—but I would like to say that one expects a responsible physicist to give some evidence, beyond mere analogy, that the physical process he suggests is real. The literature on thunderstorm electricity is cluttered up with theories which have no physical basis, and I regret that Dr. Chalmers has not waited to make the experiments he promises before rushing into print with his ideas.

References.

- (1) Simpson and Scrase, *Proc. Roy. Soc. A*, clxi. p. 309 (1939).
- (2) Simpson, *Quart. J. R. Met. Soc.* lxix. p. 1 (1942).
- (3) Rudge, *Proc. Roy. Soc. A*, xc. p. 256 (1914).
- (4) Shaw, *Proc. Phys. Soc.* xxxix. p. 449 (1927).
- (5) Simpson, *Quart. J. R. Met. Soc.* xxxviii. p. 291 (1912).
- (6) Simpson, Meteorological Office, London, *Prof. Notes* No. 82.
- (7) Simpson and Robinson, *Proc. Roy. Soc. A*, clxxvii. p. 281 (1941).

XXXIII. *Note on the Paper "The Temperature Distribution around a Spherical Hole in an Infinite Conducting Medium," by Messrs. Pugh and Harris in the Phil. Mag. p. 661 (Sept. 1942).*

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

THE problem discussed in the above-named paper can be treated shortly and easily by the Laplace Transform method. All that is needed is the solution for an Instantaneous Spherical Surface Source over $r=r_1$ at $t=0$, when radiation takes place at $r=a$ into a medium at zero temperature. This is Ex. 20, p. 236, in 'Operational Methods in Applied Mathematics' by Carslaw and Jaeger (O. U. P., 1941). The solution is given, the sign of h being different from that of the example, in Lowan's paper in the Bull. Amer. Math. Soc. vol. xxxv. p. 952 (1939), having been sent to him by me as a correction of an erroneous result in his earlier paper in that volume, p. 315.

Yours faithfully,
H. S. CARSLAW.

Burrodoo, N.S.W.,
Australia.
Dec. 12, 1942.

"We are indebted to Professor Carslaw for pointing out an erratum in equation (18) of our paper, where $\frac{Ka+k}{ka}$ has been omitted before h^2t under the exp.

H. LL. D. PUGH,
A. J. HARRIS."

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XXXIV. *On Sensory Energy, with special reference to
Vision and Colour-Vision.*

By JOHN H. SHAXBY, D.Sc., F.Inst.P.,
Physiology Institute, University College, Cardiff*.

[Received January 26, 1943.]

SUMMARY.

Transformations of the energy of a primary sensory stimulus to that of molecules of the material of a sense-organ lead to the concept of receptor energy, in the form of damped vibrations causing the ejection of electrons. It is shown that this energy, the proximal stimulus of sensory action currents, is the quantity S in the Weber-Fechner Law $S=K \log I/I_0$. The limitations of the law are discussed in connexion with the unequal sharing of stimulus energy among the receptor molecules.

The cases of steady stimuli (in the skin senses, smell, etc.) are contrasted with those of periodic stimuli (in hearing and sight). The values of the visibility function for different wave-lengths are calculated and shown to agree with the observed values.

A theory of colour-vision based on the electronic content of the discharges of action current, and not requiring the retinal triplication of the trichromatic theory, is shown to lead to an equation for a match in colour which is in accord with observed data.

I.

PHENOMENA of sensation in the various sense-organs show such striking similarities as to suggest that there must be some essential underlying general identity in their mechanisms. Thus, in all, whatever the nature of the primary external stimulus, this stimulus must attain some threshold value in order to evoke sensation; this sensation increases with increase of the stimulus; there is a just perceptible increase of sensation by a finite increase of stimulus which bears a constant ratio to the intensity of that stimulus. This relation, Weber's Law, holds good only within a certain range of stimuli well above the threshold. It was developed by Fechner into the equation $S=K \log I/I_0$, where S is a measure of the strength of sensation, I that of the stimulus, and K and I_0 are constants. The derivation of this equation has been subjected to much criticism, and the possibility of measuring sensation at all has been disputed, but there is little doubt that the formula gives a reasonable estimate of

* Communicated by the Author.

our introspective judgments of quantitative variation of sensation with change of stimulus; S , at any rate, seems to measure something recognizable in sensation which runs parallel with a logarithmic function of I .

Further, in all senses, the application of an appropriate stimulus to a sense-organ has been shown by Adrian (1928) and many others to result in action currents in the associated nerve-fibres. These currents are alike for all senses, consisting of a series of rhythmic bursts or pulses of negative electric potential, which for a given fibre under a steady stimulus occur at equal time-intervals, and are discharged by travelling along the fibre at a definite speed. They are produced only by stimuli exceeding a threshold value, and further increase of stimulus quickens up the rhythm of the discharge, but produces no change of magnitude of individual pulses (the All-or-None Law). With continued constant stimulation the rhythm gradually slows down (adaptation, in terms of sensation) and in some cases, the phasic type of receptor, the discharge ceases altogether after the initial rhythmic outburst; at cessation or marked decrease of stimulation there may be a similar burst of action pulses in these cases.

These phenomena are characteristic of the excitation of sensory receptors of all types in the many different animals studied. They clearly indicate a conversion of the energy of the external stimulus, directly or by a process of excitation, into energy of a molecular electronic or ionic nature within the receptor, energy of similar nature in all types of sense-organ which, in turn, gives rise to the electrical energy of the action currents in the nerve-fibres. It is noteworthy that, whereas the adequate stimulus for a given sense-organ is entirely ineffective for organs of other types, electrical stimulation is capable of causing their characteristic sensation in all alike.

It will be convenient to give the name Receptor Energy to this intermediate stage characterizing the activity of the molecular elements of any sensory end-organ. This paper is an investigation of the results which follow from a definite hypothesis as to the nature of the molecular processes concerned in this receptor energy. It will be seen that it leads to conclusions in strict agreement with a number of observations as to sensory phenomena which have hitherto rested on a purely empirical basis.

Among these are the phasic nature of the sensations evoked by steady stimuli as contrasted with the continuing nature of the sensations due to rhythmic stimuli (*e. g.* steady pressure on the skin, steady administration of olfactory stimuli, as compared with sound-waves reaching the ear, or light falling on the eye), Weber's Law and the Weber-Fechner equation, the nature of the failure of these relations when the stimulus is not far above its threshold, and the precise form of the visibility curve for light of different wave-lengths. It leads also to certain views as to the facts of colour-vision, which are shown to accord with known quantitative data.

The hypothesis is briefly this: the energy of the primary stimulus

sets up disturbances of some part of the receptor molecules (termed for brevity the receptor elements), and when these disturbances reach or exceed a certain threshold value the receptor elements eject an ion or electron, whose charge goes to form a unit of the total charge initiating a burst of action current. If the energy of stimulation is too small no such electrons are emitted, no action current and, therefore, no sensation results. Beyond this stage, increasing stimulus brings about increase in the number of electrons which the disturbed element ejects, and consequently gives rise to more rapid building up of the action current charge to its fixed all-or-none value, and so to a faster action current rhythm.

The disturbance can be most conveniently represented as vibrations, so that the equation describing the motion is that of an oscillation with dissipation of energy taking place as a discrete event. But the essential feature is the loss of energy in each swing of the element; the precise phase at which it occurs and the condition determining it (critical value of velocity, displacement or other feature of the oscillation) are of secondary interest to the main argument. Since a number of independent elements must be excited the phenomena can be adequately represented statistically by a drain of energy throughout an oscillation. With this simplification we can avoid equations involving discontinuities and avail ourselves of the familiar laws of an oscillation which dissipates energy against quasi-viscous resistance proportional to velocity.

The equation of such motion under zero external force is

$$\ddot{x} + 2k\dot{x} + \omega^2 x = 0,$$

and our problem is that of the effect of external forces of various types acting upon a particle under the assumed conditions.

II.

Let us first take the case of a constant stimulus, *i.e.* a steady stream of energy of amount I per unit time. This results in the application of a constant force to the element of the receptor molecule.

The final result is a displacement of the element from its initial position through a distance x_0 . When this steady state is attained both \dot{x} and \ddot{x} become zero, so, for the period during which this equilibrium is being established, the equation of motion is

$$\ddot{x} + 2k\dot{x} + \omega^2 x = \omega^2 x_0,$$

the solution of which is

$$x = be^{-kt} \sin(qt + \epsilon) + x_0,$$

where $q^2 = \omega^2 - k^2$.

Under the initial conditions $x=0$ and $\dot{x}=0$,

$$\tan \epsilon = q/k \quad \text{and} \quad b = -\omega x_0/q.$$

Hence at any instant t the displacement from the final steady position is given by

$$x_0 - x = \omega x_0/q \cdot e^{-kt} \sin(qt + \epsilon),$$

a vibration dying away exponentially at a rate determined by the dissipation coefficient k .

We have postulated that this equation holds for vibrations exceeding some critical value, no electron being expelled and therefore no energy dissipated when this limiting value is not attained.

Above this limit each vibration results in the ejection of a unit of receptor energy. Hence, if n vibrations are completed before the energy falls to the critical value, n units of receptor energy are produced, and the total receptor energy is proportional to n and to the number N of receptor elements thus excited.

The maximum displacement from the final equilibrium position in successive half-swings, *i.e.* the amplitudes of the periodic values of $x_0 - x$, decreases in the ratio $e^{-\pi k/q}$ per half-cycle, so if a_1, a_{n+1} are the amplitudes of the first and the $(n+1)$ th terms,

$$a_1/a_{n+1} = e^{n\pi k/q}.$$

If, then, a_{n+1} is that amplitude which is just too small for further emission of receptor energy, it has a constant value for the given element. On the other hand, a_1 is proportional to the steady force acting on the receptor molecule. Since energy is proportional to force multiplied by displacement, itself proportional to the force, we have

$$I/I_0 = a_1^2/a_{n+1}^2 = e^{2n\pi k/q},$$

where I is the stimulus as defined above and I_0 is a constant.

The number of vibrations exceeding the critical value $n = q/2\pi k \cdot \ln I/I_0$ and the receptor energy S is proportional to this, so $S = K \log I/I_0$, where K is a constant. This is the Weber-Fechner relation, now connecting the receptor energy S with the logarithmic function of the stimulus. We may then regard the equation as expressing the receptor energy as the immediate stimulus of sensation, which is, as estimated by introspection, proportional to it.

For the whole sense-organ this relation holds as long as K remains constant, *i.e.* as long as all the receptor elements are stimulated to sensibly equal amounts. But we cannot suppose that all elements are equally well placed for the reception of stimulus energy, so if the stimulus decreases a point will be reached at which the less favourably situated elements will be unable to contribute their share to the receptor energy, and this number will progressively increase as the stimulus falls. This has two effects, the lessening of K , already noted, and, secondly, the fact that the more favourably situated elements get a share of the external stimulating energy greater than the average causes them to continue to produce receptor energy for stimuli less than would have been the case had all the elements shared equally. This is equivalent to a lowering of I_0 .

Ranging the elements in order, the first being the least favourably placed, the N th most favourably, the total effective stimulus I

$$= i_1 + i_2 + \dots + i_N,$$

and the total receptor energy $S = s_1 + s_2 + \dots + s_N$, the subscripts denoting the individual stimuli and receptor energies respectively. If we suppose the elements identical, so that the individual threshold stimulus i_0 is the same for each, we have

$$\begin{aligned} S &= k (\log i_1/i_0 + \log i_2/i_0 + \dots + \log i_N/i_0) \\ &= k (\log \theta_1 \cdot \theta_2 \cdot \theta_3 \cdot \dots \cdot \theta_N) \\ &= kN \log \bar{\theta}, \end{aligned}$$

writing θ with the suitable subscript for the successive fractions i/i_0 , and $\bar{\theta}$ for their geometric mean.

If I is sufficiently great $\bar{\theta}$ remains practically proportional to I/I_0 , so that the Weber-Fechner relation holds. But as I decreases the less favourably situated elements begin to drop out, receiving stimuli less than the individual threshold i_0 .

In this case
$$S = k \log (\theta_m \cdot \theta_{m+1} \cdot \dots \cdot \theta_N),$$

where θ_m refers to the element just excited, the elements 1, 2, \dots ($m-1$) having fallen out.

We do not know the relation between the several θ 's, but we cannot be far wrong if we make the simplifying assumption that they can be replaced by θ 's in geometric progression, so that each bears to the preceding one a ratio r , the same throughout the series. The first θ is practically unity, since this element is barely excited.

So $S = k \log (1 \cdot r \cdot r^2 \cdot \dots \cdot r^{p-1})$, where p is the number of terms, whence $S = \frac{1}{2}kp(p-1) \log r = Cp^2$ approximately, where C is a constant, if p is not too small.

Now the number of excited elements is at any rate approximately proportional to the excess of the stimulus I above a constant value J , at which all elements just fail to be stimulated. We may thus expect that when the stimulus is decreased below the value at which it effectively excites all elements, the Weber-Fechner relation will give place to the parabolic relation $S = C(I-J)^2$, where C and J are constants. We return to the point in the special case of vision, where we have data enabling the relation to be tested.

We see that the constant I_0 of the Weber-Fechner equation is considerably larger than the absolute threshold stimulus for the sense under consideration, as is well known to be the case; stimulation is still effective below the Weber's Law range, but the Weber fraction $\Delta I/I$ becomes progressively larger as I is decreased. In other words, a just perceptible change of sensation only occurs when the corresponding finite change of stimulus ΔI bears a larger ratio to the stimulus I than was the case for those larger stimuli for which Weber's Law was valid. This at once follows from the argument above, since equal diminutions of the receptor energy S occur with a steadily diminishing number of responsive elements as the stimulus falls.

In addition to the fundamental reasons just discussed, there must be

Other factors in the failure of Weber's Law for small stimuli. The stimulus may fail to reach more deeply-seated receptors as a whole, for instance in the case of pressure exciting a number of touch receptors in the skin. This is an extreme case of the sort of thing we have considered for the elements of a single receptor; it must result in a grading of sensation analogous to the grading of response in the contraction of muscle under sub-maximal stimulation. Anatomical considerations also, the density of distribution of the particular end-organs in the region under examination, must affect the value of the sensory response to a given physical stimulus.

It should be emphasized that in the above analysis I is in all cases the effective stimulus, which is not necessarily simply proportional to the total applied external physical or chemical stimulus. When there is no qualitative difference between sensations, as for instance in the case of the series of sensations produced by lifting a graded set of weights, we may expect proportionality; the lifted weight can be taken as a measure of I in the Weber-Fechner relation. But if the sensory experiences differ in quality, *e. g.* when lights of different colours fall upon the eye, the effective stimulus I cannot be expected to be simply proportional to the incident energy; the relation which, in fact, holds in this case is discussed later in considering the visibility curve.

In the case under consideration, that of a constant stimulus, we have seen that the receptor energy is transient, so that the resulting sensation must die away during the maintenance of the stimulus. This is notably the case for the senses for which this type of stimulation occurs. The pressure of clothes upon the body or of a blunt pin held in steady contact with the skin soon ceases to be felt. The application of a steadily maintained olfactory stimulation soon fails to result in the perception of an odour. The phasic type of response is a direct consequence of the transitory nature of the receptor energy.

III.

For the senses of hearing and vision the stimuli are periodic, of the type $f \sin (nt + \alpha)$ in the simplest case, f being proportional to the amplitude of the exciting waves and $n/2\pi$ their frequency. A phase angle α at the instant $t=0$ is introduced for the sake of generality.

Instead of the steady final displacement reached by the molecular vibrators in the case of a constant stimulus already considered, we now have a permanent régime of forced vibrations, of period equal to that of the disturbing waves. During the establishment of this régime we have in general also a set of transient damped vibrations of the type considered in section II.

The equation of motion of the "permanent" forced vibrations is

$$x = a \sin (nt + \alpha - \delta),$$

where $\tan \delta = 2kn/(\omega^2 - n^2)$ and $a = f/\{(\omega^2 - n^2)^2 + 4k^2n^2\}^{\frac{1}{2}}$.

The equation for the "transients" is

$$x = be^{-kt} \sin (qt + \epsilon),$$

as before; the values of b and ϵ depend upon the initial conditions. Normally these are that $x=0$ and $\dot{x}=0$, so that writing the total displacement, the sum of the two for the permanent and transient vibrations, for $t=0$, we have

$$a \sin (\alpha - \delta) + b \sin \epsilon = 0,$$

and

$$an \cos (\alpha - \delta) + b(q \cos \epsilon - k \sin \epsilon) = 0.$$

It can be shown then that

$$\cot \epsilon = k/q - \frac{n \cot (\alpha - \delta)}{q}$$

and

$$b^2 = a^2 \sin^2 (\delta - \alpha) + a^2/q^2 \cdot \{k^2 \sin^2 (\delta - \alpha) + n^2 \cos^2 (\delta - \alpha) - 2kn \sin (\delta - \alpha) \cos (\delta - \alpha)\}.$$

Thus the displacement of the receptor element at time t can be written

$$a \sin (nt + \alpha - \delta) + be^{-kt} \sin (qt + \epsilon),$$

the values for b and ϵ being expressed in terms of a , ω , k , n and α .

With such periodic disturbances the rate of energy supply depends, not only on the applied force and the proportional amplitude of the vibrations produced, but also on the frequency of these vibrations and the dissipation coefficient k , being proportional to

$$kn^2 f^2 / \{(\omega^2 - n^2)^2 + 4k^2 n^2\}.$$

Thus, if I is the external stimulus, proportional to f^2 , then for a given type of receptor element the effective stimulus is

$$I / \left\{ \left(\frac{\omega^2}{n} - n \right)^2 + 4k^2 \right\},$$

and is thus dependent on the frequency of the incident waves.

In the case of sound-waves this frequency is extremely small compared with that of the natural vibrations of the intra-molecular receptor elements: n/ω is very small. The frequency of light-waves, on the other hand, is of the same order as that of molecular vibrations; in fact, we shall find reason to suppose that the frequency corresponding to ω is that of the light-waves for which the eye has its maximum sensitiveness, *i. e.* those for which the visibility curve attains its maximum value. The two cases must be considered separately.

Hearing. IV.

The period of a sound-wave being so long compared with that of the molecular vibrations acts as a series of steady stimuli; each wave (or half-wave) sets up a train of transients similar to those discussed in section II. above. They die away, to be renewed by the succeeding disturbance. The receptor energy is no longer of the phasic type, and sensation is consequently maintained as long as the stimulus of incident sound-waves persists.

The initial amplitude of the transients, and hence the receptor energy produced, is nearly independent of n , for a given stimulating force f ; for the expression for a above reduces to f/ω^2 when n/ω is very small, and $\tan \delta = 2kn/\omega^2$, so that δ is also small. Hence b is also nearly independent of the frequency of the sound-waves, for a given physical intensity. The actual response of the ear must depend on the number of elements disturbed by a stimulus of given pitch, and this, since notes of different pitch excite a different set of fibres of the basilar membrane, is a matter outside the scope of our inquiry.

Vision. V.

The frequency of the stimulating light-waves is of the same order as that of the molecular vibrations which they bring about. Hence the vibrations of the permanent régime (section III.) as well as those of the transients are within a range of frequencies near $\omega/2\pi$, the "natural" frequency of a receptor element. In particular, the phenomenon of resonance appears if n is sufficiently near ω .

The mechanism of the action of light upon the retina is photo-chemical; visual purple in the case of the rods, and almost certainly some allied or similar photo-sensitive substance in that of the cones, are broken down by light and are rebuilt in darkness. More generally, the processes of breaking down and building up go on simultaneously, so that under steady illumination equilibrium is attained, the number of active photo-sensitive elements remaining statistically constant. Increased illumination increases the rate of breakdown so that a new equilibrium is gradually reached, with fewer effective photo-active molecules and therefore lower sensitivity: the eye has become light-adapted. Conversely weaker light, permitting the regeneration to catch up on the breaking down, brings about dark adaptation, the increased sensitiveness to dim light.

The receptor elements considered above may or may not be parts of these photo-sensitive molecules*, but in either case it is clear that change of light intensity must affect the receptor energy, not merely for a single receptor element in the ways already discussed, but through the change in the number of effective photo-sensitive elements accompanying change in the state of adaptation. But this complication is avoided in the case of a light stimulus of fixed energy sufficient to ensure a fully light-adapted eye, and therefore an unchanging number of photo-sensitive molecules. Data of this kind are available relating sensory measurements (candle-powers, foot-candles, lamberts, etc.) with purely physical quantities (ergs, watts); thus the variation of brightness with wave-length through the spectrum under stimulation by constant energy

* We shall shortly see reason to suppose that they are, at any rate, closely associated. In any case, Occam's Razor seems to cut at the probability of the existence of two distinct kinds of molecule in the receptor cells, both sensitive in similar circumstances to the action of light-waves.

is known with some accuracy, and is represented in the so-called visibility curve.

On the views here set forward this curve exhibits the relation between receptor energy and stimulus energy, and its values enable us to check the validity of the expressions obtained for different frequencies n of the incident light.

As in the cases already dealt with the receptor energy is given by an expression of the form $K \log I/I_0$, where I is the effective stimulus to which the receptor energy is due.

In the present case I is not simply proportional to the external physical stimulus, but is also a function of the frequency of the light-waves. The magnitude of the energy dissipated per second in the maintenance of a régime of vibrations of the type of section III. above, *i.e.* that which provides the receptor energy, is given by the expression

$$\begin{aligned} & f \sin (nt + \alpha) \cdot \dot{x} \\ &= nfa \sin (nt + \alpha) \cos (nt + \alpha - \delta) \\ &= \frac{1}{2} nfa \{ \sin (2nt + 2\alpha - \delta) + \sin \delta \}. \end{aligned}$$

The first term involves no drain of energy from the stimulus, since it restores in each half-cycle what it borrowed in the preceding half. The second term, $\frac{1}{2} nfa \sin$, is independent of time, showing that the energy drain, *i.e.* the energy input into the receptor elements, is a steady one. We are thus justified in the expression $K \log I/I_0$ for the receptor energy: the situation is similar to that of the constant stimuli considered in section II. A train of transients is set up; during the time taken for the damped vibrations to subside to their critical value electrons are ejected to initiate the action currents.

But the recurring stimuli of successive light-waves maintain the ejection of electrons, so that vision persists. If the receptor elements are part of the photosensitive molecules, newly regenerated molecules present themselves and fresh phasic vibrations will be set up.

$$\frac{1}{2} nfa \sin \delta = kn^2 f^2 / \{ (\omega^2 - n^2)^2 + 4k^2 n^2 \},$$

substituting the values of a and $\sin \delta$ from section III.

Thus, for a constant energy of physical stimulation, *i.e.* a fixed value of f , the effective stimulus I is the expression just obtained, and we may write

$$S = K \log \frac{n^2}{\{ (\omega^2 - n^2)^2 + 4k^2 n^2 \} I_0} = K \log 1 / \{ (\omega^2/n - n)^2 + 4k^2 \} I_0,$$

including the constant factors k and f^2 in the constant I_0 .

$$\text{Hence} \quad e^{S/K} = I_0 \{ (\omega^2/n - n)^2 + 4k^2 \}.$$

If S_m is the maximum value of S , when $n = \omega$, $e^{-S_m/K} = I_0 \cdot 4k^2$,

$$\text{whence} \quad e^{\frac{S_m - S}{K}} = 1 + (\omega^2/n - n)^2 / 4k^2.$$

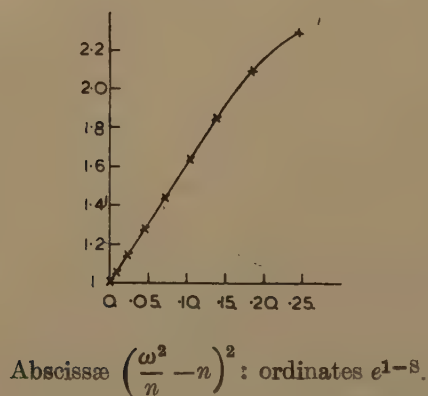
Visibility curves are usually constructed with their maximum equal to unity. By suitable choice of units we may make $K=1$ and $S_m=1$.

Then $e^{1-S} = 1 + (\omega^2/n - n)^2/4k^2$, a linear relation between the exponential and the quantity $(\omega^2/n - n)^2$.

For the light-adapted eye S_m falls at the wave-length 0.556μ . This gives the experimental value of ω , and the visibilities for various wave-lengths through the spectrum provide data for testing this linear relation. A plot of e^{1-S} against $(\omega^2/n - n)^2$, in fact, proves to be a straight line between the wave-lengths 0.500μ and 0.620μ , showing curvature for wave-lengths lying nearer either end of the spectrum, as shown in fig. 1. The slope of the linear part gives the value of $4k^2$ as 0.16 .

Adopting this value of $4k^2$ we can now calculate the visibility for any wave-length. It is to be noticed that if we interchange the terms n and ω^2/n the value of the expression containing them is unchanged, a well-known property of the expression for the energy of forced vibrations for forcing frequencies of equal energy on opposite sides of the

Fig. 1.



resonant frequency $\omega/2\pi$. That this relation holds in the case of the visibility for the range of wave-lengths specified, and even beyond it, is shown in Table I., where the wave-lengths of equal visibilities are collected in pairs, the square root of their product being given in the last column.

In the calculations which follow, advantage is taken of this relation $\lambda_1\lambda_2 = \lambda_0^2$, or $n_1n_2 = \omega^2$ (where the subscripts refer to pairs of wave-lengths or frequencies of equal visibilities, and λ_0 is the wave-length, 0.556μ , of maximum visibility), to simplify the arithmetic. We may then write n_2 instead of ω^2/n_1 , making the expression for the visibility or receptor energy S ,

$$1-S = \ln \{1 + (n_2 - n_1)^2/4k^2\}.$$

The frequencies are given simply as the reciprocals of the corresponding wave-lengths measured in μ .

Table II. shows the results of the calculations. The first two columns give the wave-lengths of the pairs of equal visibilities. The square of their frequency-differences is shown in column 3. Upon our theory the expression $1 + (n_2 - n_1)^2 / 4k^2$ should be equal to e^{1-S} , so this expression

TABLE I.

Visibility.	Wave-lengths (μ). λ_1 λ_2		$(\lambda_1 \lambda_2)^{\frac{1}{2}}$
1.000	0.556		0.556
0.952	0.542	0.570	0.556
0.870	0.533	0.580	0.556
0.757	0.524	0.590	0.556
0.631	0.5185	0.600	0.558
0.503	0.510	0.610	0.558
0.380	0.504	0.620	0.559
0.262	0.496	0.630	0.559
0.170	0.486	0.640	0.558
0.103	0.474	0.650	0.555
0.059	0.460	0.660	0.551
0.030	0.446	0.670	0.547
0.016	0.434	0.680	0.543
0.008	0.425	0.690	0.542

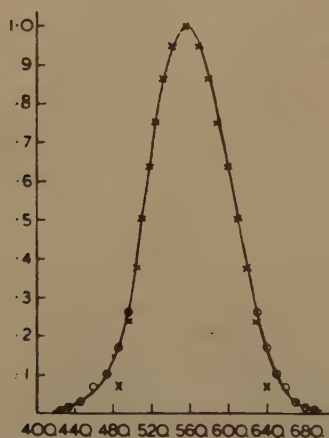
TABLE II.

1.	2.	3.	4.	5.	6.	7.
λ_1 .	λ_2 .	$(n_2 - n_1)^2$.	e^{1-S} .	S calculated.		S observed.
0.556		0	1	1		1
0.5425	0.570	0.0081	1.050	0.951	—	0.952
0.533	0.580	0.0231	1.145	0.865	—	0.870
0.524	0.590	0.0458	1.283	0.752	—	0.757
0.5185	0.600	0.0689	1.432	0.639	—	0.631
0.510	0.610	0.103	1.645	0.503	—	0.503
0.504	0.620	0.138	1.863	0.378	—	0.380
0.496	0.630	0.183	2.147	0.236	0.263	0.262
0.486	0.640	0.245	2.531	0.071	0.171	0.170
0.474	0.650	0.327	3.045	—	0.103	0.103
0.460	0.660	0.433	3.706	—	0.061	0.059
0.446	0.670	0.564	4.525	—	0.030	0.030
0.434	0.680	0.694	5.337	—	0.016	0.016
0.425	0.690	0.815	6.096	—	0.009	0.008

is given in column 4. The resulting calculated value of S, the visibility, appears in column 5. Column 6 will be explained later. Column 7 presents the experimentally determined values of S; the data of I.E.S. have been those adopted. Fig. 2 shows the observed visibility curve as a full line; the calculated values are shown by the crosses.

It will be seen that there is good agreement between observed and calculated values for wave-lengths from 0.504μ to 0.620μ , that outside these limits discrepancies appear which increase rapidly as one approaches the end of the spectrum, *i. e.* wave-lengths of low visibility, and that the last five lines of the table lead to negative values for S . This failure of the theory is a result of the factors discussed in section II. above: the stimuli for these wave-lengths are not sufficiently great, as compared with the threshold values, for the Weber Law to hold. In such cases we saw that the value of the receptor energy is given by an expression of the form $S=C(I-J)^2$, so that the relation between I and \sqrt{S} is linear. In the present case the stimulus I is the reciprocal of the expression $(n_2-n_1)^2+4k^2$, and S is the visibility. Fig. 3 exhibits the graph of the square root of the visibility plotted as ordinate against

Fig. 2.



Abcissæ wave-lengths ($m\mu$): ordinates visibilities.

I as abscissa. For completeness, and to contrast the two parts of the visibility curve for which the two distinct relations hold, the graph has been carried up to wave-lengths of higher visibilities, *i. e.* those for which the original unmodified theory is valid, the cases where the whole number N of receptor units is stimulated. For these points the curve is of course not linear.

But for the wave-lengths now to be considered, lying nearer the ends of the spectrum, it will be seen that the plotted points lie well on a straight line. (It is perhaps relevant to note here that the two wave-lengths of equal visibility are not strictly those to which the expression $(n_2-n_1)^2$ applies, as shown by the gradual changes in the value of $(\lambda_1\lambda_2)^{\frac{1}{2}}$ in the last few lines of Table I. Further, macular absorption must reduce the effective stimulus near the blue end of the spectrum. But the effects of these factors upon visibility seem to be negligible.)

The equation of the linear portion of fig. 3 is approximately

$$y=1.39(x-0.097).$$

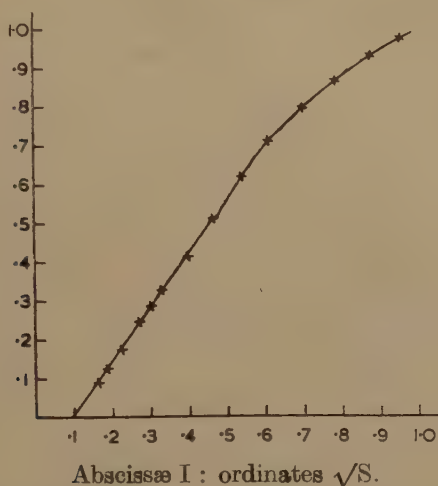
The visibilities calculated from this equation and the data for I above give the values shown in column 6 of Table II. In the visibility curve of fig. 2 these points are shown by the small circles.

VI.

The phenomena considered apply to vision in general, to the integrated activity of the whole population of retinal elements. If we examine the behaviour of single elements or small groups in the light of our hypothesis a few further conclusions may be drawn.

The activity of a single element is complicated by the fact that, under the stimulation of light, photo-sensitive material is used up, while

Fig. 3.



molecular metabolism produces fresh substance. The individual element may thus react in a phasic manner, but the precise reaction must depend, not only on the degree of exhaustion of photosensitive material, *i. e.* on the past activity of the element, but also on the phase of vibration in relation to the phase of the stimulating light-waves.

The whole picture is thus complex, but, so far as action currents are concerned, certain possibilities are indicated:—(a) action currents may continue throughout the period of illumination, (b) the initial changes occurring at the onset of illumination, producing the complex set of disturbances which precede the attainment of the steady state, cause a series of damped vibrations with a phasic ejection of electrons, and (c) elements in certain phases of vibration at the cessation of illumination will react to this cessation by a further train of damped vibrations and a further phasic ejection of electrons.

Thus the action current pattern may be one of combined activity throughout the time during which the eye is stimulated, this activity may be accompanied by initial or final phasic outbursts, or these phasic discharges may occur alone. Hartline (1938, 1940) has recorded all these phenomena in his examination of the action current responses of single retinal nerve-elements.

Colour-Vision. VII.

Colour-vision, due to cones only, is essentially a foveal phenomenon, and this central area of the retina is also that which exhibits the highest visual acuity, a result of the possession by each cone of its individual nerve-fibre, in contrast with the sharing of a nerve-fibre by a large number of rods in more peripheral regions.

Theories of colour-vision, for the most part, have postulated a triplicity of mechanism of some kind in the visual apparatus, and the chief evidence for this has lain in the facts of colour-matching. In particular, the Young-Helmholtz' theory, which it is probably true to say holds the field to-day, is, in fact, a formulation in physical terms of the colour-matching equation; it is entirely adequate in dealing with the phenomena in question.

The equation states that light of brightness c and colour C can be exactly matched by a suitable mixture of three "primary" colours; less than three are usually inadequate, more than three always unnecessary. The three primaries, if spectral (monochromatic) need to be distributed in the spectrum, one near each end, the third near the middle. In other words, the primaries are usually red, green and blue, but, with this restriction, no definite wave-length is preferable to others: we can select the wave-lengths to suit a particular problem or technical requirement.

If the brightnesses of these matching primary components are r , g and b , the colour equation is

$$c(C) = r(R) + g(G) + b(B).$$

It will be seen that this equation states two distinct facts:—(1) that $c = r + g + b$, the brightness match, requires that the brightness c is obtained by a mixture of three brightnesses whose sum is equal to c (Abney's Law); and (2) that the proportions in the mixture of R , G and B require to be those stated in order that the colour C shall be exactly matched.

It is to be noticed that, in fact, with spectral primaries, the equation never holds in the simple form stated. One of the three primaries must have a negative brightness. In other words, the match is actually produced by matching a mixture of C and the primary whose colour coefficient is negative against a mixture of the other two primaries. Thus, for instance, using the primaries $R = 0.65\mu$, $G = 0.53\mu$ and $B = 0.46\mu$, a match can be made of the colour of wave-length 0.50μ , the matching equation in terms of the units known as trichromatic coefficients being

$$1(C) = -0.233(R) + 0.772(G) + 0.461(B).$$

That is to say, a certain amount of the red primary has to be added to the light of wave-length 0.50μ , and this mixture can then be matched by a mixture of green and blue primaries. The equation, in fact, is

$$1(0.50) + 0.233(R) = 0.772(G) + 0.461(B).$$

In terms of real monochromatic mixtures, then, the triplicity of the colour-matching equation largely disappears.

If triple mechanisms exist, where are they situated? It is to be emphasized, as has been done by Guild and others, that the differentiation must exist at both ends of the chain, in both retina and visual cortex. We must suppose that there is some differentiation in the cone-layer which permits discrimination in its action according to the wave-length of the incident light. This discrimination, once set up, must be preserved at the cerebral cortex, since it appears as the sensory phenomenon of colour difference. This involves a further point, that the differences must exist also in the action currents in fibres of the optic nerve and tract. The only escape from this conclusion is to suppose the difference to be topographical, for instance, in the existence of three types of cones, associated with their own nerve-fibres and with corresponding localization at the cortical end by which the three types of cones produce recognizably different sensations. But segregated cortical areas cannot exist for this purpose: the topography of the visual cortex is concerned with the excellent local sign of the sense of sight.

There is a further objection: visual acuity is as good with monochromatic light as with white, under the same conditions of brightness. Whatever the precise view we adopt to explain the limits of visual acuity, it is clear that the spacing of the cones (or rods) must play a dominating part. But if there are three kinds of cones, and, for instance, the "red" cones only are active when the stimulating light is near the long-wave end of the spectrum, then only a fraction of the order of one-third of the number of cones available for white light is brought into play. Under such circumstances there would inevitably be a marked diminution of acuity as compared with that for white light.

Another possibility is that the triplicity exists in each cone, in three photo-sensitive substances or at any rate three photo-chemical processes, for red, green and blue respectively, each acting independently of and without interaction with the others. Such a view does not remove the difficulty as to the threefold differentiation of the processes in nerve-fibres and cortex. There is a considerable and growing body of evidence that there are several distinct photo-sensitive substances in the retinas of many animals, and some evidence that they may be derived from or closely allied to visual purple as a sort of parent substance. Recent work by Ragnar Granit (1941) and his colleagues has attacked the problem from a different angle, the elementary action currents of various animal eyes, dark-adapted or partially or fully light-adapted, under the stimulus of lights of different wave-lengths. Small elements of the retina are tested by a technique of micro-electrodes, and the

action currents examined by the cathode-ray screen or loud-speaker after suitable amplification. The author considers that his findings support Young's classic theory that there are fibres in the retina carrying impulses excited by different regions of the spectrum. The action current response to different wave-lengths is shown to vary for different points in the retina. The micro-electrode detects elements in the frog with narrow maxima, of a "red" type, about 0.60μ , a "green" type, about $0.54-0.56\mu$, and a "blue" type, about 0.48μ . The aggregate of all these elements results, with micro-electrodes of "very large" cross-section, in a curve of broader maximum with its peak about 0.56μ . Granit finds no cases in which a unit delivers a message of only one type; the part of the spectrum to which response is maximal depends on the state of adaptation of the eye. While this work makes a strong case for a polychromatic response of the retinal elements for the frog and other animals, it is much less clear that the variability of response with the element examined (and with its state of adaptation) is three-fold. Different "green" elements show marked variations among themselves, both in their general form of response to light of different wave-lengths and in the positions of their maxima; the same is true of "red" and "blue" elements.

It seems fair to conclude that light-adaptation is accompanied by (perhaps is due to) the production of a definite number of distinct photo-sensitive substances, each showing a characteristic absorption spectrum and response to light, and that these materials exhibit their individual features in the action current responses of single elements (or groups of a few elements). The integrated response is determined (and this is Granit's opinion) by the relative numbers of the various types, *i. e.*, the relative proportions of the several photo-products; it therefore changes with the state of adaptation of the eye. The change in the position of the maximum and to some extent in the form of the human visibility curve may probably be regarded as the same process regarded from the sensory aspect.

But such a polychromatic response of retinal elements (or even a strictly trichromatic one, if later work should show that the varied responses really fall into three distinct classes) in no way lessens the difficulty of the three types of action current demanded by the trichromatic theory of vision. This difference of behaviour of the nerve-fibre under stimulation of the retina by lights of different wave-lengths has clearly to be faced. Some progressive change with changing wave-length is thinkable, but to demand three clear-cut types of response only increases the difficulties. Change of rhythm cannot be invoked, for this is a function of the intensity of stimulus, and is the basis of perception of differences of brightness.

In fact, there seems very little evidence for triplicity at all; trichromatic theories seem to depend upon the triplicity suggested by the colour-matching equation, and as we have seen this is not convincingly triple. The fact that light of a given colour can be matched by a mixture

of three other lights is no proof of a triplicate mechanism in the eye, in the absence of other evidence: a mixture of water and alcohol of density equal to that of petroleum could be identified with the latter if we had no other data than those of density by which to differentiate them.

The visibility curve, almost symmetrical about its maximum, gives no hint of being built up of three distinct curves, as has been pointed out by Houstoun and others. The extreme asymmetry and difference in size of the three primary constituent curves obtained from the colour equations do nothing to increase belief that the visibility curve is the sum of the three in any physical sense. Further, we have seen above that the form and the exact values of the visibility curve can be deduced from the hypothesis of receptor energy, with no differentiation between the various parts of the spectrum, except the ordinary gradual change of frequency from one to the other.

Of course, the practical convenience and importance of the analytical treatment of colour by the colour-matching equation is not in question. It is of the highest value in practical problems of colour description and matching: but these are quite other matters than the mechanisms of the eye.

VIII.

If we regard the processes of colour-vision as essentially single in nature, with a progressive change of some sort as we pass along the spectrum, we introduce no fresh difficulties, but on the contrary remove the need for a thrice repeated triplicity, in retina, nerve-fibre and cortex, so difficult to conceive in living tissues.

Brightness relations have been dealt with above in considering the visibility curve. In a complete colour match, then, the matching of brightness implies that the total receptor energy must be the same for the matching mixture as for the matched colour. There must be some further condition to account for the match in colour, some second equality between the two streams of action current.

Action currents consist of series of discrete discharges, of definite frequency for a single fibre under a given stimulus. Increase of stimulus does not change the potential difference associated with each of these discharges, but quickens up the rhythm, so that the total energy delivered per second is increased; a stronger sensation results.

For a brightness match these deliveries per second must be equal; difference in colour, with equal brightness, can then only be due to a difference between the individual discrete discharges making up the two streams. Equality of periodicity is already assured by the equality of the two currents, since the all-or-none law holds. For a colour match there must be some equality between the packets making up the separate discharges, and this equality may be either in the number of quanta giving rise to the charge associated with each packet, or, more probably, in the electronic energy making up the packet.

For matched intensities the periodicity of discharge is the same for the two streams, so the quotient of the energy-flow per second divided by the quantum energy, $h\nu$, is proportional to the number of quanta associated with the production of each packet of action current. We have seen above that luminosity is the measure of receptor energy: thus we must divide the luminosity of each component of a match by $h\nu$, where ν is the frequency of the component. The quotient for a matched colour should then be equal to the sum of the component quotients for the matching colours, if the first hypothesis is the true one.

But, since with an intensity match the energy in each packet is the same for the matched lights, it would appear necessary that the energy of the electrons composing the packets should be the same for both, rather than their number. Now in the expulsion of an electron a certain amount of energy is lost in freeing the electron from its molecular constraints. If this energy is c , we have $h\nu - c$ as the energy of the expelled electron. The efficiency of the process is thus $(h\nu - c)/h\nu$. Each term of the quotients of luminosity divided by $h\nu$ must then be further multiplied by its appropriate efficiency, as thus expressed, to obtain the relative energies due to the same number of electrons, in order to build up the same energy for different wave-lengths of incident light.

Thus, if S represents the luminosity of any term in a colour-match, of frequency ν , then the term representing the relative numerical contribution to the action current packet is $S/h\nu$, which is proportional to $S\lambda$, where λ is the wave-length.

To obtain the equality of the energy contribution we must multiply this by $(h\nu - c)/h\nu$, or by $(1 - k\lambda)$, where k is a constant.

Thus, if light of wave-length λ and luminosity S_λ is completely matched by a mixture of three other wave-lengths, λ_r , λ_g and λ_b , of luminosities S_r , S_g and S_b respectively, the ordinary equation for equality of brightness is

$$S_\lambda = S_r + S_g + S_b, \quad (1)$$

while the equation representing the identity of colour is either, on the first hypothesis,

$$S_\lambda \lambda = S_r \lambda_r + S_g \lambda_g + S_b \lambda_b,$$

or, on the second hypothesis,

$$S_\lambda \lambda (1 - k\lambda) = S_r \lambda_r (1 - k\lambda_r) + S_g \lambda_g (1 - k\lambda_g) + S_b \lambda_b (1 - k\lambda_b). \quad . . . (2)$$

The pair of equations (1) and (2) give an exact specification of the quantities and wave-lengths concerned in a complete colour-match, and their solution is unique. Thus, if the first equation is varied by changing the wave-length of one constituent, keeping the energy unaltered, then the second equation is no longer satisfied, since the change of wave-length involves a change in the number of quanta. Similarly, if equality in the second equation is maintained by substituting for one constituent an appropriate number of quanta of a

different wave-length, there can no longer be equality of receptor energy, or therefore of brightness.

A typical set of colour-matches in terms of trichromatic coefficients is provided by the accurately observed data of W. D. Wright in his Report upon "A Re-Determination of the Trichromatic Mixture Data" (Wright, 1929). Dr. Wright has been good enough to furnish me with the factors for recalculating these matches in terms of sensation values, *i. e.*, of the relative luminosities of the constituents of the matches.

TABLE III.

Wave-length (μ).	Component Luminosities.			Total Luminosity. S_{λ} .
	Red S_r .	Green S_g .	Blue S_b .	
0.41	0.025	-0.058	0.0539	0.021
0.43	0.022	-0.053	0.0540	0.023
0.45	0.010	-0.030	0.0543	0.034
0.46 (Primary)	—	—	0.0541	0.054
0.47	-0.015	0.070	0.0537	0.109
0.48	-0.046	0.224	0.0493	0.227
0.49	-0.082	0.517	0.0406	0.476
0.50	-0.113	0.950	0.0250	0.862
0.51	-0.101	1.232	0.0111	1.142
0.52	-0.054	1.290	0.0033	1.239
0.53 (Primary)	—	1.230	—	1.230
0.54	0.060	1.109	-0.0012	1.168
0.55	0.113	0.980	-0.0016	1.091
0.56	0.172	0.831	-0.0017	1.001
0.57	0.233	0.668	-0.0012	0.900
0.58	0.294	0.509	-0.0009	0.802
0.59	0.350	0.357	-0.0005	0.706
0.60	0.394	0.241	-0.0004	0.635
0.61	0.428	0.153	-0.0003	0.581
0.62	0.451	0.092	-0.0002	0.543
0.63	0.470	0.046	-0.0002	0.516
0.64	0.480	0.017	-0.0001	0.497
0.65 (Primary)	0.486	—	—	0.486
0.68	0.491	-0.014	—	0.477
0.70	0.493	-0.018	—	0.475

The primaries used in Dr. Wright's Report are "real," viz., the spectral colours of wave-lengths 0.65μ , 0.53μ and 0.46μ , and Table I. of his Report specifies the matches for 25 spectral colours ranging from 0.41μ to 0.70μ in terms of trichromatic coefficients based on these primaries.

To obtain from these coefficients the luminosities (*i. e.* the receptor energies) we must proceed as follows:—The mean luminosities of the three primaries are $L_r=0.486$, $L_g=1.23$ and $L_b=0.054$ (data supplied by Dr. Wright). To express the colour-matches in terms of luminosities we must multiply the three trichromatic coefficients of Table I. of the

Report by these factors respectively. Thus, for instance, for wave-length 0.60μ , the coefficients are $R=0.811$, $G=0.196$ and $B=-0.007$. Multiplying each of these by the appropriate conversion factor quoted, we find the values 0.394 , 0.241 and -0.00038 for the relative luminosities of the constituent primaries, matching a total luminosity of $0.394+0.241-0.00038=0.635$ of the matched wave-length 0.60μ . This equation is, of course, the first equation of the defining pair, the intensity equation.

This recalculation has been made for each of the 25 wave-lengths of Table I. of Dr. Wright's Report, and Table III. (see preceding page) is the restatement of that Table in terms of component and total luminosities, instead of trichromatic coefficients. These figures, denoted as S_r , S_g , S_b and S_λ respectively, are then proportional to the receptor energies concerned.

Table IV. gives the result of applying the theory above to these data. Column 1 gives the wave-length, 2, 3 and 4 the products $S_r\lambda_r$, $S_g\lambda_g$ and $S_b\lambda_b$, 5 the sum of these quantities, 6 the product $S_\lambda\lambda$. Thus column 5 consists of values proportional to the number of quanta associated with the mixture, or to the number of ejected electrons: column 6 gives the corresponding number for the matched wave-length. It will be seen that the pairs of columns 5 and 6 are nearly equal, but with considerable discrepancies near the ends of the spectrum.

In columns 7, 8 and 9 are the products obtained by multiplying the figures of columns 2, 3 and 4 by the appropriate factor $(1-k\lambda_r)$, $(1-k\lambda_g)$ and $(1-k\lambda_b)$ respectively. The value $k=0.75$ has been used in this computation. The quantities in these three columns, Q_r , Q_g and Q_b are thus proportional to the respective electronic energies. Column 10 gives their sum, and column 11 the corresponding product Q_λ of the data in column 6 multiplied by $(1-k\lambda)$ for the matched wave-lengths. The pairs of data in these columns agree closely throughout.

The value $k=0.75$ corresponds to $c=1.47\times 10^{-12}$. (The quantum for wave-length 0.65μ is 3.03×10^{-12} , that for wave-length 0.46μ is 4.28×10^{-12} .)

We thus see that the conditions for an exact match of two illuminations are:—

(1) The intensity relation: the receptor energies of the matched lights must be equal; and

(2) The colour relation: the individual packets of discharge of the two action currents must possess equal electronic energies.

The data of Table IV. show that this colour relation is in excellent agreement with the observed data for colour-matches.

It then appears that the determining factor for colour is the quantum of light waves, as determining the charge of the individual packets of action current, and that the variation due to the change of magnitude of the quantum as we pass along the spectrum affords an adequate physical basis for the phenomenon of change of colour, without recourse

TABLE IV.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
Wave-length (μ).	$S_r\lambda_r$	$S_g\lambda_g$	$S_b\lambda_b$	Sum.	$S\lambda$	Q_r	Q_g	Q_b	Sum.	Q_λ
0.41	0.0162	-0.0307	0.0248	0.0103	0.0086	0.0083	-0.0185	0.0162	0.0060	0.0060
0.43	0.0143	-0.0281	0.0248	0.0110	0.0099	0.0073	-0.0169	0.0163	0.0067	0.0067
0.45	0.0065	-0.0159	0.0250	0.0156	0.0154	0.0033	-0.0096	0.0164	0.0101	0.0101
0.46 (Primary)	—	—	0.0248	—	—	—	—	0.0163	—	—
0.47	-0.0098	0.0371	0.0247	0.0520	0.0512	-0.0050	0.0223	0.0162	0.0335	0.0332
0.48	-0.0299	0.1187	0.0227	0.1115	0.1091	-0.0153	0.0715	0.0149	0.0711	0.0698
0.49	-0.0533	0.2740	0.0185	0.2392	0.2332	-0.0273	0.165	0.0122	0.150	0.148
0.50	-0.0735	0.5035	0.0115	0.4415	0.4310	-0.0376	0.303	0.0075	0.273	0.270
0.51	-0.0657	0.6530	0.0051	0.5924	0.5824	-0.0336	0.393	0.0033	0.363	0.360
0.52	-0.0351	0.6837	0.0015	0.6501	0.6440	-0.0180	0.412	0.0010	0.395	0.393
0.53 (Primary)	—	0.6519	—	—	—	—	0.393	—	—	—
0.54	0.0390	0.5878	-0.0005	0.6263	0.6307	0.0200	0.354	-0.0004	0.374	0.375
0.55	0.0735	0.5194	-0.0007	0.5922	0.6000	0.0376	0.313	-0.0005	0.350	0.352
0.56	0.1116	0.4404	-0.0008	0.5512	0.5606	0.0573	0.265	-0.0005	0.322	0.325
0.57	0.1515	0.3540	-0.0005	0.5050	0.5130	0.0776	0.213	-0.0004	0.290	0.294
0.58	0.1911	0.2677	-0.0004	0.4584	0.4652	0.0979	0.162	-0.0003	0.260	0.263
0.59	0.2275	0.1892	-0.0002	0.4165	0.4165	0.117	0.114	-0.0001	0.231	0.232
0.60	0.2561	0.1277	-0.0002	0.3836	0.3810	0.131	0.077	-0.0001	0.208	0.209
0.61	0.2782	0.0811	-0.0001	0.3592	0.3544	0.143	0.049	-0.0001	0.192	0.192
0.62	0.2934	0.0487	-0.0001	0.3420	0.3367	0.150	0.029	-0.0001	0.179	0.180
0.63	0.3055	0.0244	-0.0001	0.3298	0.3251	0.156	0.015	-0.0001	0.171	0.171
0.64	0.3120	0.0090	-0.0001	0.3209	0.3141	0.160	0.005	—	0.165	0.165
0.65 (Primary)	0.3159	—	—	—	—	0.162	—	—	—	—
0.68	0.3192	-0.0074	—	0.3118	0.3244	0.163	-0.004	—	0.159	0.159
0.70	0.3205	-0.0095	—	0.3110	0.3325	0.164	-0.009	—	0.155	0.158

to sharply specialized retinal mechanisms concerned with three primary colours.

IX.

The physical conditions for a colour match have just been obtained, but the question of what defines a colour cannot be so simply answered. The distinction between, say, red and blue depends on the fact that the equal energies of the action current packets arise from fewer quanta for blue than for red, but the sensations of blueness or redness can no more be explained in purely physical terms than can the difference between the sensations concerned in hearing and sight.

A colour-match is obtained when the two equations (1) and (2) above are satisfied : but the converse, that a given value of the quantity $\lambda(1-k\lambda)$ should define a colour is not necessarily true. In fact, we can easily see that the same value of this quantity can refer to two quite different colours. A notable instance is that of white light. Since this can be produced by a suitable mixture of the three primaries, it is clear that the mean value of $\lambda(1-k\lambda)$ for the mixture must correspond to that for some particular wave-length, *i. e.*, to light of some definite spectral colour. The mean value is obtained by determining $S\lambda(1-k\lambda)$ for each primary and dividing the sum of the three by the total intensity, *i. e.*, by the expression $\Sigma\{S\lambda(1-k\lambda)\}/\Sigma S$.

Using Dr. Wright's Table I., converted to luminosities as above, we find the three *S*'s are 0.486, 1.230 and 0.0541, while the three terms of the numerator are 0.162, 0.393 and 0.0162, whence the desired expression is 0.323.

Any pair of complementary colours can similarly be used for the calculation. The result is necessarily the same as that just obtained, apart from experimental errors in the data. For instance, the complementary wave-lengths 0.60μ and 0.48μ give the value 0.320, and the pairs 0.585μ and 0.485μ the value 0.323.

We can also calculate the desired quantity by integrating the values of $S\lambda(1-k\lambda)$ throughout the white light spectrum. This can be done approximately by adding the values for equal small steps of λ , and dividing the sum by the corresponding sum for the luminosities. Dr. Wright's spectrum is that for a black body of 5000°K ., and, using the visibility curve and the black-body radiation for this temperature, for steps of 0.01μ from 0.40μ to 0.70μ , we find that the above quotient is 0.323.

This is the value of $\lambda(1-k\lambda)$ for a wave-length of about $0.54\text{--}0.55\mu$, *i. e.*, for a spectral green. It must also be the value for a suitable mixture of the two primaries 0.65μ and 0.46μ , *i. e.*, for a purple, the colour complementary to the above green. Hence the value 0.323, while it must be given for a match of any of the three "colours," white, green and purple, does not characterize the colour unambiguously.

So, in general, the quantity $\lambda(1-k\lambda)$ for any wave-length can be duplicated by a suitable mixture of two primaries, although, as we have seen, such a binary mixture cannot yield a colour-match. In fact, to

match any spectral colour by the admixture of two primaries we must add a certain amount of the third primary to the colour. We may express this otherwise; the given colour, together with a certain amount of white light, can be matched by a mixture of two primaries. For instance, referring to Table III. we see that the wave-length 0.50μ is matched if a certain amount of the red primary is added to it by a mixture of the green and blue primaries. In terms of trichromatic coefficients (Wright's Table I.) the equation of the match is

$$1(0.50) + 0.233 \text{ Red} = 0.772 \text{ (Green)} + 0.461 \text{ (Blue)}.$$

This can be written

$$1(0.50) + 0.699 \text{ (White)} = 1.005 \text{ (Green)} + 0.694 \text{ (Blue)}.$$

In other words, a mixture of two spectral primaries matches a monochromatic colour if the latter is diluted with white. The appropriate proportions of the components depend, of course, upon the choice of spectral primaries, but the fact that the matching mixture of the appropriate two primaries produces a tint less saturated than the light of the matched monochromatic colour is perfectly general.

This, then, may give the clue to the physical condition controlling an observed colour sensation. We cannot give a definition of complete saturation in colour perception: we can always imagine a more saturated tint than any observed one. In experimental fact, if the eye is exposed for some time to light of any colour and then observes the complementary colour, the latter appears more saturated than it does under normal observation. Any observed spectral tint, then, presents itself with a certain non-definable degree of saturation, and an attempt to match it by two spectral primaries (*i. e.*, any two other monochromatic lights) of appropriate positions in the spectrum, results in a less saturated version of the given tint. We may surmise that the colour observed is governed by the wave-lengths of the components making it up, and that greater spread of these components, adding a white component, lessens saturation. In terms of action currents then, the position is this: light of a single wave-length causes action current packets which all arise from quanta of the same magnitude, and therefore contain the same number of electrons. A matching light which is a mixture of two wave-lengths gives rise to action current packets which are only equal to those of the former case as an average, containing packets which contain fewer than the average number (derived from the component of shorter wave-length) and others with more than the average (from the longer wave-length). The resulting sensation then will be intermediate between those which would be produced by these longer and shorter waves acting alone; the mean effect will correspond with that of the original matched monochromatic colour, but there will be a tendency towards the effect which would be produced if lights of all wave-lengths were used, *i. e.*, towards the sensation of whiteness. The greater the difference between the wave-lengths of the matching binary mixture the greater will be

this tendency. An exact colour-match can only be restored if we add to the original monochromatic light a stimulus appropriate to produce the same spread in the make-up of the action current packets, viz., a certain quantity of the third primary component or of white light. To sum up, the view is put forward that the observed colour depends on the number of electrons required to make up the action current packet of the fixed (all-or-none) energy content, and so, ultimately, on the magnitude of the stimulating light quanta: homogeneous packets, *i. e.* those which all contain the same number of quanta, produce a tint of greater saturation than heterogeneous packets, *i. e.* those of the same average electron content but with a spread of more or less varied values about the given mean number. Complete spread with the relative intensities of components, which exist in the white light spectrum, give the sensation of whiteness.

Equality of brightness depends on the reception by the cortex of action currents of equal energy per second, which is equivalent to an equality of the frequencies of the action current rhythm, since the energy per packet is fixed, under the all-or-none rule. Further, the phenomena of colour depend upon the make-up of the individual packets. A colour-match results if the average electric charge associated with a packet is the same for the matched beams, and if the spread of the individual packets about that average is the same for both. If the average value is the same, but the spread different, the more homogeneous case produces a sensation of a more saturated tint of the same colour. Difference of tint results from a difference in the average value of the charge of the individual action current packets, the tint being the more saturated as the variation between the charges making up the series of packets is smaller.

X.

The hypothesis put forward above covers the main facts of colour-mixture and colour-matching without recourse to any sharply marked triplicity of structures anywhere in the optical system of the eye. It does not, however, demand an absolute similarity of all retinal elements; as mentioned above, the work of Granit supports the view that there are differences in the mode of response of individual receptor elements under the stimulus of lights of differing wave-lengths and different states of adaptation of the eye. This diversity implies that the integrated activity of the eye produces phenomena which are averages; for example, the photopic visibility curve must be a composite of a series of narrower curves with somewhat differing points along the spectrum for their maxima. This conception of differences in sensitivity of different elements to different wave-lengths, variations distributed about a mean but not depending in any way on the existence of three distinct sharply differentiated classes of receptor elements, fits perfectly well into the general view of this paper.

Further, it would seem to afford an explanation of some, at any rate, of the phenomena of simultaneous contrast and after-images. For instance, if the eye is exposed for some time to light of a given colour, the elements whose maximum of response lies nearest to that colour will undergo the processes of adaptation more strongly than those whose maxima of response are further removed. Subsequent stimulation by white light will then cause a maximum stimulation in the less light-adapted elements, and so give rise to a sensation tending towards the previously unstimulated wave-lengths, *i. e.*, to that of the complementary colour. Again, if the second stimulation is by a monochromatic light of a different colour, the constituents appropriate to the former stimulation will be relatively weakened; in terms of action current, the spread of the variation of charge between different packets will be reduced, and the colour will therefore appear more nearly saturated than it normally does.

Granit has suggested that phenomena of inhibition play a part in the phenomenon of simultaneous contrast: the probability of this is increased by the dependence of colour upon the action currents, and not merely upon the retinal response. The depression with stimulation by a given wave-length falling upon part of the retina, if carried over in this way to neighbouring areas, will result again in the preferential activity of those receptors whose maxima lie furthest from the stimulating light, and will cause relatively greater numbers of action current packets set up by these receptors, and so the appearance of the complementary colour.

ADDENDUM.

It was remarked in the opening section that the ordinary equations of damped vibrations were adopted for mathematical convenience and as giving the type of motion consistent with loss of energy to the ejected electron in each cycle of the intra-molecular vibrations; the actual event, an impulsive loss of energy at the moment of ejection, at some particular phase of the motion, it was unnecessary to specify more exactly for the general hypothesis. It may be of interest, however, to consider one possible way of ejection, impact with perfect elasticity between the vibrating particle and the electron. The particle loses a certain amount of energy to the electron, which is thereupon ejected if the velocity it acquires is sufficiently great. Such an impact recurs in each cycle, and if m is the mass of the particle and μ that of the electron, the ordinary laws of impact show that

$$(m-\mu)/(m+\mu)=v_1/v_0=v_2/v_1\ldots=v_{n+1}/v_n \doteq r, \text{ say,}$$

where v_0, v_1, v_2 , etc., represent velocities of the particle in the successive stages. Similarly, if V_1, V_2 , etc., represent the velocities acquired by the series of electrons,

$$V_2/V_1=V_3/V_2=\ldots=V_{n+1}/V_n=r.$$

The energy of the vibrating particle is thus reduced abruptly in the ratio r^2 in each cycle, and its amplitude of vibration in the ratio r . Hence its vibrations are of the nature already considered, the quantity r replacing the term $e^{-2n\pi k/\omega}$ for each single vibration.

Similarly, the velocity which successive electrons attain becomes less and less, and the ejection ceases when V_{n+1} becomes too small for the electron to overcome its molecular constraints and to escape.

ACKNOWLEDGMENTS.

I desire to express my thanks to Professor T. Graham Brown, F.R.S., and to Mr. G. F. Herrenden Harker for their helpful criticism and suggestions, and to Dr. W. D. Wright for his courtesy in affording data for the calculations for the section on Colour-Vision.

REFERENCES.

- Adrian, 'The Basis of Sensation' (1928): also many papers in Journ. Physiol.
 Granit, J. O. S. A. xxxi. p. 570 (1941).
 Hartline, Am. Journ. Physiol. cxxi. p. 400 (1938); J. O. S. A. xxx. p. 239 (1940).
 Wright, "A Re-determination of the Trichromatic Mixture Data" (Medical Research Council: Report VII. of the Committee upon the Physiology of Vision) (1929).

XXXV. *The Evaluation of the Complex Roots of Algebraic Equations.*

By A. F. CORNOCK, M.Sc., and JOAN M. HUGHES.

(Communication from the Staff of the Research Laboratories of
 The General Electric Company, Limited, England*.)

[Received November 2, 1942.]

SUMMARY.

A method is described for the evaluation of the complex roots of algebraic equations in which the roots are evaluated directly from the original equation by a process somewhat akin to Horner's method for real roots.

1. *Introduction.*

THERE appear to be three effective methods available at the present time for the evaluation of the complex roots of algebraic equations:

- (a) The Root-Squaring method as modified by Brodetsky and Smeal ⁽¹⁾.
- (b) The Test Function method of Frazer and Duncan ⁽²⁾.
- (c) The modified Bernoulli method due to Aitken ⁽³⁾.

* Communicated by C. C. Paterson, F.R.S.

When a root has been located by any of these methods, the accuracy may be improved (if desired) by the use of Bairstow's process ⁽⁴⁾, and it appears to be largely a matter of taste for the computer to decide whether to locate a root to the required accuracy directly by the use of the process by which it was first located, or to locate it only approximately in the first instance and then to improve it by the use of Bairstow's method. Of the three primary processes for locating roots, that of Brodetsky and Smeal appears to be the most expeditious; and if any alternative process is to be of interest, it should at least not take appreciably longer than does the modified Root-Squaring process, and it should possess some virtue of its own.

The present process appears to us to have the virtue that there is no great temptation to make errors of sign (which seem to us to be the most frequent errors when developing the coefficients of the equations having squared roots), and, if made, such errors are not difficult to detect. On the other hand, it locates only a single complex root (or pair of conjugate roots) at once, while the Root-Squaring process prepares the equation for the location of all the roots. Whether others will find it to be as economical of time as its competing processes, is a matter which must be left to the experience of computers.

2. Description of the Process.

The process consists, in principle, in choosing a suitable point on the real axis to which to transfer the origin, and then, using Horner's process, to transfer the origin in the direction of the imaginary axis to a point close enough to the root for the resulting equation to be readily solvable by iteration.

Consider the equation

$$f(z) = \phi + i\psi \\ = z^n + a_1 z^{n-1} + \dots + a_n = 0, \quad \dots \quad (1)$$

where the coefficients a_1, \dots, a_n , are all real and it is supposed that n is even and that all real roots have been removed. Writing $z = x + iy$ and separating the real and imaginary parts,

$$\phi = \phi_0 + \phi_2 y^2 + \phi_4 y^4 + \dots + y^n, \quad \dots \quad (2)$$

$$\psi = \phi_1 y + \phi_3 y^3 + \dots + \phi_{n-1} y^{n-1}, \quad \dots \quad (3)$$

where the ϕ_r 's are functions of x . The general nature of the shapes of the surfaces ϕ and ψ is not difficult to visualize.

In the first place, ϕ and ψ satisfy

$$\nabla^2 \phi = \nabla^2 \psi = 0 \quad \dots \quad (4)$$

(since they are the real and imaginary parts respectively of the analytic function f). But

$$\nabla^2 w = 0$$

is the equation giving the deflexion w (normal to the x, y plane) of a soap bubble, or other flexible membrane under tension, which is free

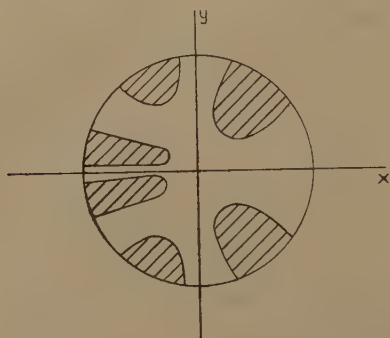
from load on its surface ; and the shape of the surface can be visualized from this physical analogy if the deflexion is known along a closed boundary. A convenient boundary for the present purpose is a circle having its centre at the origin ; and writing

$$z=r(\cos \theta+i \sin \theta),$$

$$\phi=r^n \cos n\theta+a_1r^{n-1} \cos (n-1)\theta+\ldots+a_n,$$

$$\psi=r^n \sin n\theta+a_1r^{n-1} \sin (n-1)\theta+\ldots+a_{n-1}r \sin \theta,$$

so that, for a sufficiently large value of r , ϕ and ψ are each represented by a surface having in it a number of flutes or valleys which run inwards towards the origin. The ϕ -surface is everywhere positive along the real axis and is an even function of y and θ ; and therefore, within the circular boundary, it somewhat resembles the top of a cake with a number of roughly radial valleys in its surface. The contour lines $\phi=0$ might be (for $n=6$) somewhat as sketched in the figure, the areas where ϕ is negative being shaded.



The ψ -surface has the real axis as a line of zero height, and it also is a fluted surface at a sufficient distance from the origin ; but whereas the ϕ -surface was an even function of θ , the ψ -surface is an odd function of θ , and its form is therefore more difficult to visualize. Perhaps the best analogy is to consider it as having been produced by taking a strip of rubber, laying it along the x -axis and twisting it clockwise at some points, anti-clockwise at others, and then superposing flutes so that they are anti-symmetrical with respect to the x -axis.

The roots of the equation are, of course, the points at which the contours $\phi=0$ of the valleys of the ϕ -surface intersect the contours $\psi=0$ of the flutes of the ψ -surface. A root can be approximated to by iteration once a way has been found into that valley in the ϕ -surface which contains that root. From the physical analogy it is obvious that a suitable point of departure is a point at which there is a minimum of ϕ (as a function of x) on the real axis, since such a point is (or is near) the head of a valley of the ϕ -surface. Calling α the value of x for which ϕ is a minimum, one accordingly proceeds to evaluate $f(\alpha+i\beta)$ for a series of values of β and to find a point for which $|f|^2$ is a minimum.

$f(z)$ is then transformed to $(\alpha+i\beta)$ as origin, and the process can then generally be completed by iteration.

Analytically, the argument from the physical analogy is easily justified. At a minimum of ϕ on the real axis, $\partial^2\phi/dx^2$ is positive, and $\partial\phi/dx$ vanishes. $\partial\phi/dy$ also vanishes. But if $\partial^2\phi/dx^2$ is positive, $\partial^2\phi/dy^2$ is negative—from (4)—while, on the real axis, $d^2\psi/dy^2$ vanishes, by (3). Accordingly, if ϕ_0 be the value of ϕ at the minimum on the real axis, the value at $\alpha+i\beta$ is (if β be small)

$$\begin{aligned}\phi &= \phi_0 + \left(\frac{\partial\phi}{\partial x} \delta x + \frac{\partial\phi}{\partial y} \delta y \right) + \frac{1}{2!} \left\{ \frac{\partial^2\phi}{\partial x^2} (\delta x)^2 + 2 \frac{\partial^2\phi}{\partial x \partial y} (\delta x \delta y) + \frac{\partial^2\phi}{\partial y^2} (\delta y)^2 \right\}, \\ &= \phi_0 + \frac{1}{2!} \frac{\partial^2\phi}{\partial y^2} \beta^2,\end{aligned}$$

which is less than ϕ_0 . To the same order of small quantities,

$$\psi = \psi_0 = 0.$$

Accordingly, $|f(\alpha+i\beta)|^2 < |f(\alpha)|^2$, if β is small enough. The conditions for convergence to the root during the remainder of the process could be treated similarly, but analytical elaboration does not seem worth while in connexion with a numerical process.

3. Numerical Examples.

As a simple example, one may take the equation used by Fraser and Duncan (*loc. cit.*, page 71).

$$x^6 + 3x^5 + 20x^4 + 3x^3 + 2x^2 + 4x + 5 = 0. \quad . \quad . \quad . \quad (5)$$

Differentiating, and equating to zero,

$$6x^5 + 15x^4 + 80x^3 + 9x^2 + 4x + 4 = 0, \quad . \quad . \quad . \quad (6)$$

and a minimum is immediately found at about $x = -0.5$. Reducing the root by this quantity by Horner's process gives the new equation

$$z^6 + 16.25z^4 - 32z^3 + 24.6875z^2 - 5z + 4.296875 = 0, \quad . \quad . \quad (7)$$

for which $|f|^2 = (4.296875)^2 \doteq 18.46$ at the origin.

In principle it is now convenient to evaluate for $i\beta = 0.1i$ and $i\beta = 1.0i$ —since this involves very little numerical work, and examination of the resulting values of $|f|^2$ at once indicates whether there is probably a minimum less than (or near to) $0.1i$, or between $0.1i$ and $1.0i$, and also indicates the next trial value of β . (Sometimes, of course, evaluation at $0.1i$ shows that it is unnecessary to evaluate at $1.0i$ also). In the present instance,

$$\beta = 0.1.$$

1	0	16.25	-32	24.6875	-5	4.296875
0.1i	-0.01	+1.624i	-0.1624	-3.2i	0.32 + 2.45251i	-0.245251 - 0.468i
<u>0.1i</u>	<u>16.24</u>	<u>-32 + 1.624i</u>	<u>24.5251 - 3.2i</u>	<u>4.68 + 2.45251i</u>	<u>4.051624 - 0.468i</u>	

$$|f|^2 \doteq 16.64.$$

Evaluating similarly for $\beta = 1$ gives $|f|^2 \doteq 755.43$. Clearly, the minimum lies nearer $\beta = 0.1$ than $\beta = 1.0$. Evaluation at $\beta = 0.2$ gives

$|f|^2=11.51$, which suggests evaluation at $\beta=0.4$. It is clear from the result (given below) that the absolute value of the term independent of z is so small that the root can now be found by iteration: the table for the equation having the reduced values of the roots is accordingly completed, as follows:—

1 0	16.25	-32	24.6875	-5	4.296875
0.4i	-0.16	+ 6.436i	-2.5744-12.8i	5.12+ 8.84524i	-3.538096+0.048i
0.4i	16.09	-32+ 6.436i	22.1131-12.8i	0.12+ 8.84524i	0.758779+0.048i
0.4i	-0.32	+ 6.308i	-5.0976-12.8i	10.24+ 6.80620i	
0.8i	15.77	-32+12.744i	17.0155-25.6i	10.36+15.65144i	
0.4i	-0.48	6.116i	-7.5440-12.8i		
1.2i	15.29	-32+18.860i	9.4715-38.4i		
0.4i	-0.64	5.860i			
1.6i	14.65	-32+24.720i			
0.4i	-0.80				
2.0i	13.85				
0.4i					
2.4i					

The equation for z , when the original root has been reduced by $-0.5+0.4i$ is therefore

$$z^6+2.4iz^5+13.85z^4+(-32+24.72i)z^3 \\ + (9.4715-38.4i)z^2+(10.36+15.65144i)z \\ +0.758779+0.048i=0.$$

Rearranging,

$$z=-(2.44457 \times 10^{-2}-3.22982 \times 10^{-2}i) \\ +(1.42746+1.55002i)z^2+(-1.57207 \times 10^{-1}-2.148599i)z^3 \\ +(\text{terms of higher order in } z),$$

and this can be solved by iteration exactly as in the case where a real root is to be found. The result is

$$x=-0.52259+0.42922i$$

with a possible error of about a unit in the last place.

As a second example, consider the equation

$$x^8-x^7-11x^6+10x^5+56x^4+62x^3+247x^2+214x+158=0 \quad . \quad . \quad (8)$$

given on page 87 of Brodetsky and Smeal's paper. A minimum of ϕ exists in the neighbourhood of $x=-0.6$. Reduction of the root by -0.6 gives

$$z^8-5.8z^7+3.28z^6+30.136z^5-16.768z^4+2.22944z^3 \\ +216.315328z^2-52.762381z+111.139674 \\ =0,$$

for which $|f|^2=12.352$.

Evaluation at $\beta=0.1$ and $\beta=1$ gave 11907 and 15784.6, respectively, for $|f|^2$. Accordingly, evaluations were then carried out at $\beta=0.5$, 0.7 and 0.8 giving $|f|^2=3790$, 1034, and 2256, respectively. The table

for the coefficients of the equation was therefore completed for $\beta=0.7$, giving the equation

$$\begin{aligned} z^8 + (-5.8 + 5.9i)z^7 - (12.70 + 28.4i)z^6 + (89.762 - 7.637i)z^5 \\ + (-22.0110 + 175.0462i)z^4 - (194.14882 + 58.95883i)z^3 \\ + (273.829470 - 49.150080i)z^2 + (-15.084841 + 328.460644i)z \\ + 0.790926 - 32.155752i = 0. \end{aligned} \quad (9)$$

At this point it is necessary to decide whether to complete the solution by iteration, or to make a further reduction of the root. Since the squares of the absolute values of the term independent of z , the term in z , and the term in z^2 , are 1034, 108000, and 88900, respectively, it was decided that direct iteration would be laborious and that a further reduction of the root was preferable. Writing the equation as

$$z^8 + \dots + a_7z + a_8 = 0 \quad (10)$$

the reduction made is $\delta z = -a_8/a_7$,
i. e., in this case

$$\begin{aligned} \delta z &= \frac{0.790926 - 32.155752i}{15.084841 - 328.460644i} \\ &= 0.0980089 - 0.0020881i. \end{aligned} \quad (11)$$

The total amount by which the root has been reduced is now

$$\begin{aligned} -0.6 + 0.7i + \delta z \\ = -0.501991 + 0.697912i. \end{aligned} \quad (12)$$

It was then noticed that an error had been made in the reduction to equation (9)—as appears from inspection of the imaginary part of the coefficient of z^7 . Accordingly the roots of equation (8) were reduced *de novo* by the amount (12)—this being done in two stages, first reduction by the real part, then reduction by the imaginary part, to minimise the chance of mistakes. For economy in paper, the resulting tables are not reproduced. They look somewhat fearsome, owing to the number of figures retained, but Horner's method is so well suited to use with a calculating machine that the work involved is not so serious as it seems. The final result is

$$\begin{aligned} z^8 + (-5.015928 + 5.583296i)z^7 - (14.068475 + 24.504732i)z^6 \\ + (82.061996 - 20.838089i)z^5 + (17.944374 + 167.002376i)z^4 \\ + (-192.879661 + 7.155125i)z^3 + (216.801604 - 124.973798i)z^2 \\ + (32.276674 + 303.285817i)z + (2.373895 - 1.239148i) \\ = 0. \end{aligned}$$

Only the last three terms are needed for iteration, giving

$$z = 3.2163 \times 10^{-3} + 8.1695 \times 10^{-3}i + (0.332227 + 0.750199i)z^2$$

whence, finally,

$$\begin{aligned} z &= 3.1581 \times 10^{-3} + 8.1447 \times 10^{-3}i, \text{ and} \\ x &= -0.498833 + 0.706057i, \end{aligned}$$

giving the corresponding quadratic factor

$$x^2 + 0.997666x + 0.747351. \quad . \quad . \quad . \quad . \quad (13)$$

To test the accuracy of this factor, Bairstow's process was used. The remainder after dividing (8) by (13) was

$$-3.35 \times 10^{-4}x - 3.12 \times 10^{-4} = R_1x + R_0$$

whilst the second remainder was

$$-35.820045x + 198.825886 = R_2x + R_1.$$

The corrections to the coefficient of x and to the constant term in the quadratic factor are accordingly 6×10^{-6} and 5×10^{-6} , respectively, so that the factor becomes

$$x^2 + 0.997672x + 0.747356,$$

the alteration being just about of the order of the residual accuracy of the calculation.

After any one quadratic factor has been found, it is divided out, and the next factor found similarly.

4. *Comments.*

It will be seen that, once a valley of the ϕ -function has been entered, the work involved in improving each stage of the approximation is, so long as the Horner process of reduction is being used, about twice that required in evaluating a real root by Horner's method—which is to be expected, since both real and imaginary parts have to be found. On the whole, it seems that the Root-Squaring method will be more advantageous when the absolute values of the roots differ substantially from one another (apart from questions of errors creeping in), and the present method may have its uses when the absolute values of the roots approximate to one another.

One other remark may not be inappropriate. The reduction of the type (11) transfers the origin in the direction of maximum negative slope of $|f|^2$ at the origin to which the equation was referred before the reduction, and is, in the complex plane, the equivalent of Newton's method of approximation for real roots.

Bibliography.

- (1) Brodetsky and Smeal, Proc. Camb. Phil. Soc. vol. xxii. p. 83.
- (2) Frazer and Duncan, Proc. Roy. Soc. (A), vol. cxxv. p. 68 (1929).
- (3) Aitken, Proc. Roy. Soc. Edinburgh, vol. xlv. p. 289 (1926).
- (4) See Frazer and Duncan, *loc. cit.* p. 80.

XXXVI. *The Absorption of Infra-red Radiation by Water-Vapour and Carbon-Dioxide.*

By M. McCAIG, Ph.D., M.Sc., A.Inst.P., University of Manchester *.

[Received January 7, 1943.]

The experiments described below were carried out in the Heat Transmission Department of the City and Guilds Engineering College, and the paper is an abridged form of a "Thesis approved for the Degree of Doctor of Philosophy in the University of London."

1. INTRODUCTION.

(a) *Historical.*

THE influence of pressure on the absorption of infra-red radiation by gases has been known to physicists since the end of last century. First observed by Paschen ⁽¹⁾ in 1894, it was later studied by Ångström ⁽²⁾, ⁽³⁾, Eva v. Bahr ⁽⁴⁾, and others ⁽⁵⁾, ⁽⁶⁾, ⁽⁷⁾, ⁽⁸⁾. In general, it was found that if the product pL of the pressure p and length of absorbing column L were kept constant, the fraction of the incident radiation absorbed increased with increase of p , and thus contravened Beer's law. The addition of a gas such as dry air or nitrogen, which does not itself absorb infra-red radiation, also increased the fraction of the incident radiation absorbed. According to most of the early investigators, the same, or nearly the same increase in absorption was caused by a given increase of pressure, whether produced by compressing the absorbing gas, or by adding a foreign gas. If this result were true, it would follow that for a constant total pressure, the absorption by a mixture of an absorbing and a non-absorbing gas would depend on the value of pL , and not on the individual values of p and L , p being the partial pressure of the absorbing component, and thus for this special case Beer's law would hold.

In the course of the above investigations, considerable data on the absorption of infra-red radiation by gases was collected. Much of this data had been obtained for rather badly defined sources of radiation, such as a Bunsen burner, and to calculate with its aid the total radiation emitted or absorbed, for any specified values of the temperature, path-length, and pressure, required much computational work. These computations were undertaken by Schack ⁽⁹⁾, who was largely responsible for pointing out the important contribution made by gas radiation to

* Communicated by the Author.

heat transmission in some industrial plant. The computations were so complex, and the data so incomplete, that some simplifying assumptions of dubious validity were necessary. Consequently, measurements of a type that could be applied more directly to the calculation of heat transmission were made, notably by Schmidt⁽¹⁰⁾, who measured the emission of radiation by steam; and by Hottel and Mangelsdorf⁽¹¹⁾, who measured the emission due to mixtures of steam and air, carbon dioxide and air, and steam and carbon dioxide, as well as making some measurements of the absorption of black body radiation by these gases. All these measurements were made with the total pressure atmospheric, and therefore approximately constant. When Beer's law was used to calculate the emissivity of mixtures of water-vapour and air, using Schmidt's values for pure steam, some of the results obtained were 75 per cent. greater than Hottel's measurements. In 1937 Schmidt and Eckert⁽¹²⁾, published some results of the emission of radiation by mixtures of water-vapour and nitrogen, which indicated that even with the total pressure constant, the actual value of p as well as of the product pL was important; this result, they suggested, was the explanation of the earlier discrepancies. The present work was undertaken as a consequence of this paper; during its progress a further paper by Eckert⁽¹³⁾ appeared, giving correction factors by which measurements made on the emission of radiation by pure steam can be reduced to give values for mixtures of steam and nitrogen at the same total pressure and with the same value of $p_w L$, p_w being the partial pressure of water-vapour. The corrections are large, being as much as 60 per cent. when $p_w = 0.025$ atmosphere. No such corrections were found necessary for mixtures of carbon dioxide and nitrogen.

The aims of the present work were:—

- (a) to obtain heat transmission data for conditions not previously investigated.
- (b) to check Eckert's proposed corrections.
- (c) to investigate the reasons for the departures from Beer's law.

With these ends in view, the percentage absorption of radiation from a small electric furnace, by water-vapour, carbon dioxide, and mixtures of these gases with each other and with nitrogen, was measured. The gases were contained in tubes approximately 10 in., 20 in. and 40 in. long, and the total pressure as well as the partial pressures of the individual gases could be varied from less than 0.1 to more than 2 atmospheres. The author has aimed at expressing his results in a form convenient for heat transfer calculations.

(b) *Theoretical.*

The near infra-red spectrum of a gas consists of broad bands each of which corresponds to a particular change in the vibrational energy of the molecules. If these bands are sufficiently resolved, they are found

to be made up of numerous lines, each of which corresponds to a particular change in the rotational energy of the molecules. The width of lines is well known to be increased by pressure, although most of the investigations had until recently been made in the visible or ultra-violet regions. In general terms pressure broadening is due to the perturbations of molecular energy levels by forces exerted by neighbouring molecules. Thus, if a molecule undergoes a transition from one energy level to another, the frequency of the radiation quantum emitted or absorbed is affected by the presence of these neighbouring molecules.

If monochromatic radiation of intensity I_0 passes through an absorbing gas of thickness L , and pressure p , the intensity is reduced to a value

$$I = I_0 e^{-kpL}, \quad \dots \dots \dots (1)$$

where k is the absorption coefficient. The half-width δ of a spectral line is defined as the difference in wave-numbers between the points on either side of the line at which

$$k = \frac{1}{2}k_m, \quad \dots \dots \dots (2)$$

where k_m is the maximum value of the absorption coefficient at the centre of the line. The half-width so defined is not the same as the distance between the points at which the absorption is half the maximum value, as this quantity unlike δ increases with L .

The theory of the broadening of spectral lines has been surveyed by Margenau and Watson⁽¹⁴⁾, Elsasser⁽¹⁵⁾ and Schulz⁽¹⁶⁾. Even at the lowest pressures a spectral line has a certain natural width in addition to its Doppler width, but for most normal pressures the width is mainly due to pressure broadening. The effect of pressure on δ can be calculated by the statistical theory which considers the probability of a molecule being perturbed to any given extent by all its neighbours, or by the Lorentz theory which supposes all the perturbations to occur during collisions, the duration of which is short compared with the intervals between them. While the statistical theory predicts unsymmetrical lines with the half-width proportional to the square of the pressure, the collision theory predicts symmetrical lines with δ proportional to the pressure. Except at very high pressures, experimental evidence shows that the half-width is proportional to the first power of the pressure, as predicted by the Lorentz collision theory. The simplest form of this collision theory gives for the distribution of absorption within a spectral line

$$k_\nu = \frac{c\delta/2}{(\nu - \nu_0)^2 + (\delta/2)^2}, \quad \dots \dots \dots (3)$$

where k_ν is the absorption coefficient for wave-number ν , ν_0 is the wave-number of the centre of the line, and c is a constant.

Consider the variation in the total absorption due to a spectral line when p is increased, keeping pL constant. δ is increased and therefore, from equation (3), the absorption coefficient k_ν is decreased near the centre of the line, and increased at the edges of the line. For a line of

small intensity the net change in the total absorption is found from equation (1) to be small, but for an intense line the absorption near the centre remains 100 per cent. even after k_v has been reduced. Hence the increase in k_v at the edges of the line increases the total absorption. Both theory and experiment indicate that in many cases δ increases more rapidly with an increase in the pressure of the absorbing gas (resonance broadening) than with the addition of a non-absorbing gas. For such a substance Beer's law must break down even when the total pressure is kept constant. There is reason to believe that water-vapour is such a substance, since the water-vapour molecule possesses a permanent electric moment, and therefore the perturbing forces responsible for pressure broadening are likely to be greater between two water-vapour molecules than between molecules of water-vapour and a non-absorbing gas. Cornell⁽¹⁷⁾ at first found little difference between the resonance broadening of water-vapour and that due to a non-absorbing gas, but in a later paper⁽¹⁸⁾ he gave values for δ of 0.65 cm.^{-1} per atmosphere in the case of resonance broadening, and 0.4 cm.^{-1} per atmosphere in the case of broadening due to a foreign gas. These considerations suggest that Beer's law should break down for mixtures of water-vapour and a non-absorbing gas even when the total pressure is constant. Actually the use of Beer's law, which has frequently been made in such cases, is not justified by any experimental evidence, and is probably due to a mistaken analogy with the results for carbon-dioxide and non-absorbing gases.

This brief theoretical account has been given in terms of absorption, but the general conclusions hold for emission also.

2. APPARATUS.

The apparatus is shown diagrammatically in fig. 1.

The source of radiation was a small cylindrical electric furnace, one end of which was closed by a fireclay plug. Radiation, emerging through a small opening at the other end of the furnace, passed through a hole in a screen, which could be closed by means of a shutter. After traversing the absorption tube, the radiation was finally received by a thermopile.

The absorption tubes, of internal diameter 1 in., were made of brass $\frac{1}{4}$ in. thick. Both ends of the tubes were closed by rocksalt windows, rubber washers being inserted between windows and tubes. Springs (not shown in the diagram) pressed the windows evenly on the washers. To prevent cold air from reaching the windows electric heaters were wound on brass cylinders, which projected beyond the ends of the absorption tubes.

Each absorption tube had two side tubes, one leading either through flasks containing anhydrous calcium chloride to a rotary vacuum pump, or directly to a supply of nitrogen or carbon dioxide, and the other leading to the gauge. Both side tubes could be closed by means of improvised

stop valves, consisting of rubber pressure tubing and a screw pinch-cock.

The gauge also served as a source of water-vapour. It was essentially a U-tube manometer containing mercury, but had certain unusual features. While the mercury surface in contact with the atmosphere moved up and down a narrow-bore glass tube about three metres long, the other surface moved less than one centimetre in a steel cylinder about five centimetres in diameter. A little water floated on the surface of the mercury in this cylinder, which was surrounded by an oil bath maintained at any desired temperature by an electrical thermostatic system. This temperature was kept lower than that of any part of the absorption or connecting tubes, and so it determined the pressure of water-vapour in the system. The total pressure in the absorption tube could be calculated, knowing the height of the mercury surface in the glass tube, provided the height of this surface when the gauge was evacuated or at atmospheric pressure was also known. Allowance was made for the movement of the mercury surface in the steel cylinder

Fig. 1.

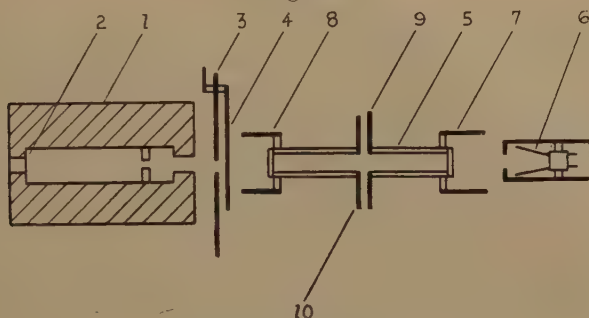


Diagram of apparatus.

1. Furnace. 2. Fireclay plug. 3. Screen. 4. Shutter. 5. Absorption tube.
6. Thermopile. 7, 8. Heated cylinders protecting windows. 9, 10. Side tubes.

3. EXPERIMENTAL PROCEDURE.

Readings of the thermopile, and of the thermocouples which gave the temperatures of the absorption tube and furnace, were made with a Tinsley potentiometer, graduated to one microvolt.

When the temperatures of the various parts of the apparatus had become steady, the absorption tube was evacuated with the valve leading to the gauge closed. With the pump still running, this valve was then opened for two or three periods of a few seconds each, in order to allow air and other gases to be swept from the gauge by the evaporating water. Three successive readings of the thermopile were then taken (*a*) with the shutter open, (*b*) with the shutter closed, and (*c*) with the shutter open again.

$$D = (a+c)/2 - b \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

was then a measure of the radiation reaching the thermopile from the furnace. The valve on the pump lead was closed and that on the gauge pipe opened. Water in the gauge evaporated and filled the absorption tube with vapour at the desired pressure.

Thermopile readings were then taken in the same way as before, giving

$$D' = (\alpha' + c')/2 - b'. \quad (5)$$

To investigate mixtures of water-vapour and other gases, the pump was disconnected, and the new gas added to the water-vapour in the tube. After allowing time for the new gas and water-vapour to mix sufficiently, as shown by the criterion that no further change in the absorption occurred, the pressure and thermopile readings were again taken. Usually two or three concentrations of the foreign gas were investigated in one run, before the absorption tube was again evacuated, and the value of D checked. An experiment with rotating sectors showed that the e.m.f. of the thermopile was proportional to the radiation falling on it; hence the percentage absorption could be calculated from the equation

$$A = 100(D - D')/D. \quad (6)$$

The furnace temperature was noted at frequent intervals during the run, so that the values of D in this equation could be corrected for any fluctuations; the time of each observation was also noted.

To test whether any error was introduced by water-vapour being swept out of the absorption tube into the gauge when the second gas was introduced, the effect of keeping the gauge valve closed during this operation was tried. The results agreed with those obtained with the valve open.

Before making measurements on carbon dioxide alone or mixed with nitrogen, the gauge was dried. The procedure was similar to that adopted with water-vapour, except that in some cases the gases were mixed before being admitted to the absorption tube.

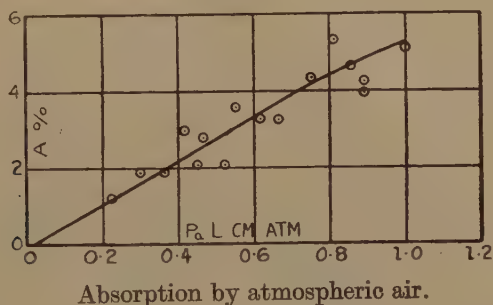
Some of the radiation is absorbed by water-vapour and carbon dioxide in the air outside the absorption tube. Hence the radiation is deficient in those wave-lengths which would be most strongly absorbed by the gases inside the tube, and the measured absorption is too low. To enable this correction to be calculated, experiments were made in which atmospheric air was admitted into the absorption tube. In fig. 2 the resulting absorption is plotted against $p_a L$, where p_a is the vapour pressure of water in the atmosphere, and L is the length of the absorption tube.

The dependence of the absorption on p_a suggested that it was due mainly to water-vapour, and this hypothesis was confirmed by extrapolating the results for mixtures of water-vapour and nitrogen (to be described later) to low values of the water-vapour pressure. This assumption does not necessarily conflict with the more important role attributed to carbon dioxide by Gerlach⁽¹⁹⁾, since his results were for radiation of longer wave-lengths. The error is largest when the absorbing

gas outside the absorption tube is the same as that inside. Thus, assuming all the atmospheric absorption to be due to water-vapour, a correction was made only when the gas inside the tube was also water-vapour. If some of the atmospheric absorption was actually due to carbon dioxide, the correction made would be too high. As the corrections themselves were small this source of error cannot be very important.

The estimation of the correction involved certain rather arbitrary assumptions. If the absorption within the more important water-vapour bands were 100 per cent. and continuous, about half the radiation from a black-body source at $1,000^{\circ}\text{C}$. would be absorbed. From the appearance of these bands under moderate resolution, they were assumed to consist of lines of equal intensity spaced 18 cm.^{-1} apart. Higher resolution reveals many intermediate lines of smaller intensity, but fortunately the precise nature of the assumptions does not greatly affect the final value of the correction; *e. g.* it is scarcely altered if twice the number

Fig. 2.



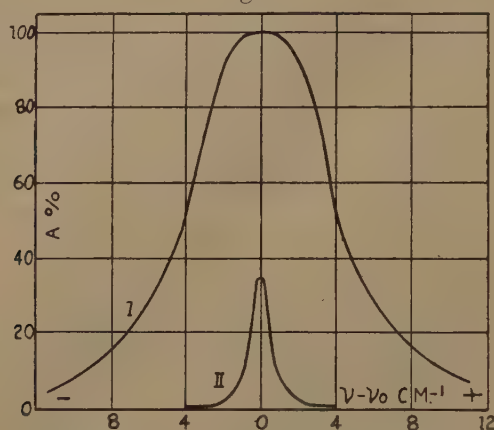
of lines each having half the intensity is assumed. Taking equation 3 to represent the contour of the lines, the value of the constant c which gave the actual value of the absorption for a number of different combinations of the pressure and path-length was determined. It was possible to find one value of c which predicted the absorption fairly accurately for a number of widely different values of the pressures and path-lengths, although for other values divergences up to 30 per cent. did occur. Using this value of c , the contour of a spectral line according to equation (3) was plotted for the pressures and path-lengths both inside and outside the absorption tube (fig. 3). The required correction was calculated for various frequencies within a spectral line, using these two curves. The mean correction for the line was computed graphically, and taken as typical for the whole spectrum.

The correction was found to be about the same for all combinations of the partial pressures of water-vapour and nitrogen, and of the path-length which gave rise to the same absorption. Hence in fig. 4 the estimated correction is plotted against the measured absorption for several values of p_a .

A small absorption occurred when nitrogen was admitted into the absorption tube. Being small, it was not noticed until the longest tube was used. At first the absorption was attributed to some impurity in the nitrogen, but a statement received from the suppliers made this explanation appear unlikely, and soon afterwards the investigation had to be suspended.

The pump would not reduce the pressure in the absorption tube below 3 mm., but as this pressure of gas, even if it were all water-vapour,

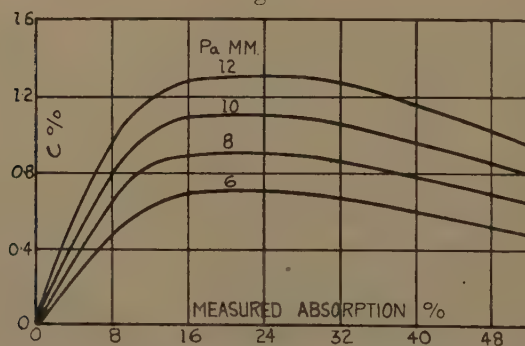
Fig. 3.



Distribution of absorption in spectral line.

- I. Curve for steam at atmospheric pressure.
 II. Curve for air containing 10 mm. of mercury of water-vapour.
 Path-length = 25.4 cm.

Fig. 4.

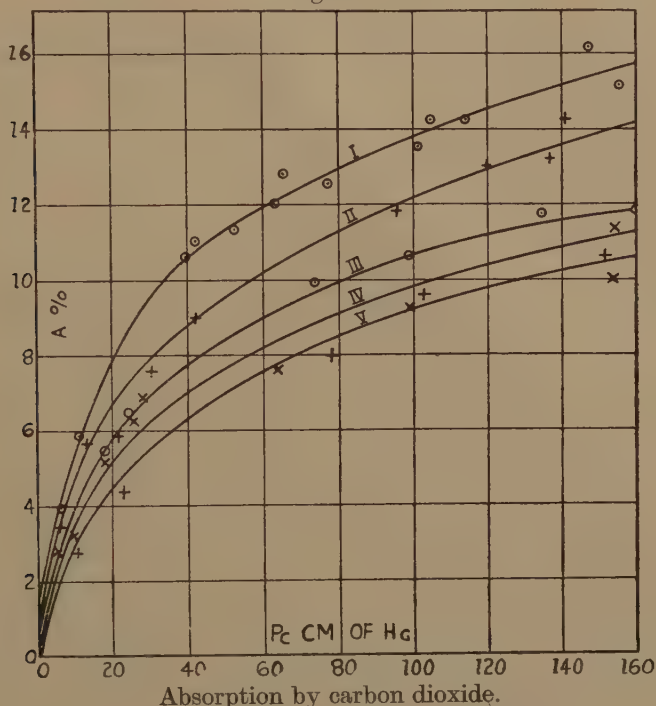


Correction for absorption by 25.4 cm. path in air outside tube.

would not absorb appreciably, no action had been taken. It was later noticed that the addition of nitrogen to this quantity of water-vapour would increase the absorption by about the amount observed, 1.3 per cent. for one atmosphere of nitrogen in the longest tube. As the residual gas not removed by the pump may well have been water-vapour, the

consequences of this explanation were considered. No appreciable error would be produced in the results for water-vapour, but the absorption for a given pressure of carbon dioxide would be too high by an amount rather less than the absorption, observed with the same pressure of nitrogen. In reaching this conclusion it is assumed that carbon dioxide increases the water-vapour absorption just as nitrogen does, but the water-vapour and carbon-dioxide bands overlap slightly. It was considered that results having the least possible error would be obtained by subtracting from the carbon-dioxide results half the absorption produced by the same quantity of nitrogen. This procedure has been adopted in drawing the curves for fig. 7.

Fig. 5.



Absorption by carbon dioxide.

I....	T=139° C.	L=102.4 cm.
II....	T=134° C.	L= 50.8 „
III....	T=165° C.	L= 25.7 „
IV....	T=104° C.	L= 25.7 „
V....	T= 27° C.	L= 25.7 „

In comparing the results with those of other workers, the effective spectral distribution of the radiation was assumed to be that of a black or grey body. This assumption requires that there shall neither be selective absorption by the windows or selective sensitivity of the thermopile, and that the furnace shall emit black or grey body radiation.

To test the combined effect of all these factors, the variation of the thermopile reading with furnace temperature was investigated. Small

departures from Stefan's law were observed. It was not possible to deduce exactly what departures from black-body distribution corresponded to the observations, but, making the most unfavourable assumptions, the observed absorption would be 7 per cent. higher than

TABLE I.

p_c cm. of Hg.	p_n	A. Per cent.	p_c cm. of Hg.	p_n	A. Per cent.
L=25.7 cm.					
1.8	4.8	1.2	6.5	2.9	3.1
4.2	10.5	2.8	14.2	6.4	4.7
6.2	15.6	4.5	21.9	9.9	5.3
12.7	31.8	6.1	41.3	18.7	7.9
15.5	39.1	5.7	43.4	19.7	7.4
22.7	57.1	7.1	81.8	37.0	10.0
31.6	79.5	8.0	94.1	42.8	10.3
36.5	92.1	9.4	—	—	—
L=50.8 cm.					
1.7	3.1	1.8	5.2	2.2	3.5
2.5	4.5	2.4	15.4	6.5	5.8
11.5	18.0	6.3	40.5	17.0	8.9
14.9	26.6	6.8	76.2	32.0	10.7
25.3	45.5	9.1	93.7	39.3	11.6
29.5	52.8	8.8	110.7	46.3	12.0
45.0	81.3	10.1	—	—	—
51.3	92.5	11.3	4.8	8.6	4.8
L=102.4 cm.					
19.8	—	8.2	20.7	32.7	9.8
19.8	47.6	10.1	20.7	127.6	10.8
19.8	66.2	10.0	3.8	—	3.7
19.8	104.1	10.5	3.8	10.8	5.1
1.2	—	0.3	3.8	82.3	6.8
1.2	27.6	1.4	—	50.7	1.1
1.2	47.1	1.8	15.9	50.7	9.5
1.2	151.2	3.2	26.3	50.7	10.5
20.7	—	8.6	38.2	50.7	11.3
20.7	15.8	9.2	59.3	50.7	12.7
—	—	—	79.7	50.7	13.9

for black-body radiation. Thus, if the measured absorption were 10 per cent., the true value for black-body radiation might be as low as 9.3 per cent., although a more probable value would be about 9.6 or 9.7 per cent.

The validity of conclusions on the way in which the absorption depends on the total and partial pressures is unaffected by this source of error.

4. RESULTS.

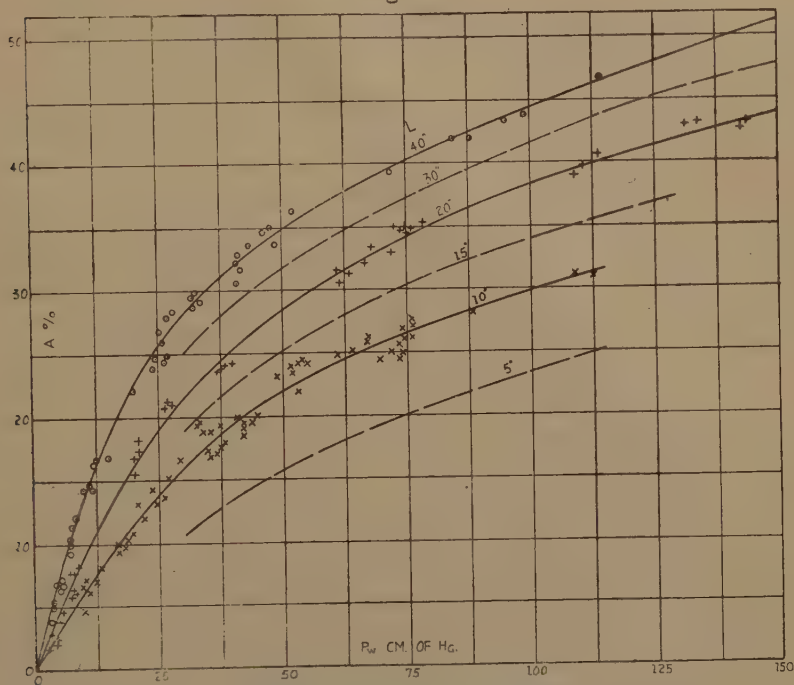
Unless otherwise stated the temperature of the furnace was about $1,000^{\circ}\text{C}$. Fig. 5 shows the percentage absorption for carbon dioxide alone.

Table I. gives the percentage absorption for mixtures of carbon dioxide and nitrogen. Wherever, as in this table, the temperature of the gas is not stated, it may be taken as 135°C .

Fig. 6 shows the percentage absorption for water-vapour alone.

As a large number of observations were made on the absorption by mixtures of water-vapour and nitrogen, a smaller number of results which have been read off from a graph are shown in Table II.

Fig. 6.



Absorption by water-vapour.
Broken lines are interpolated.

The results for mixtures of water-vapour and carbon dioxide can be most conveniently shown in a rather indirect manner, and are therefore left until the discussion.

5. DISCUSSION.

(a) Carbon-Dioxide and Mixtures of Carbon-Dioxide and Nitrogen.

For mixtures of carbon dioxide and nitrogen the facts are well established and fairly simple, and a comparison of the results with those obtained by previous workers provides a check on the accuracy of the

experimental method. Hottel and Mangelsdorf⁽¹¹⁾ used only one path-length and one total pressure, but for this limited range their results for a furnace temperature of 1000° C. and a gas temperature of 135° C.

TABLE II.

Absorption by water-vapour and nitrogen.

P_w cm. of Hg.	0	10	P_n 20	50	100	150 cm. of Hg.
L=25.7 cm.						
10	6.4	7.8	8.4	9.4	10.4	11.6
20	11.6	12.3	12.9	14.1	15.4	—
30	15.9	—	—	—	—	—
40	19.5	20.0	20.3	20.9	21.4	—
50	22.0	—	—	—	—	—
75	26.5	—	27.2	27.6	28.0	—
110	31.0	—	—	—	—	—
L=50.8 cm.						
7.5	6.8	8.5	9.1	11.0	12.7	—
20	16.1	16.9	17.6	19.8	20.8	21.3
40	26.3	26.9	27.3	28.1	29.3	—
70	33.1	33.6	34.0	34.8	35.1	—
140	43.0	—	—	—	—	—
L=102.4 cm.						
5	7.1	8.4	9.1	12.1	—	—
10	13.5	14.7	15.6	18.5	20.0	—
25	22.4	22.9	23.2	24.1	24.9	—
50	34.7	—	—	35.8	36.7	—
100	44.4	—	—	—	—	—
170	53.5	—	—	—	—	—

TABLE III.

L. cm.	$P_{c/l.c.}$ cm. of Hg.	P_n cm. of Hg.	Percentage Hottel.	Absorption McCaig.
50.8	76.0	—	10.6	10.8
50.8	38.0	38.0	8.9	9.5
50.8	6.5	69.5	5.2	5.2

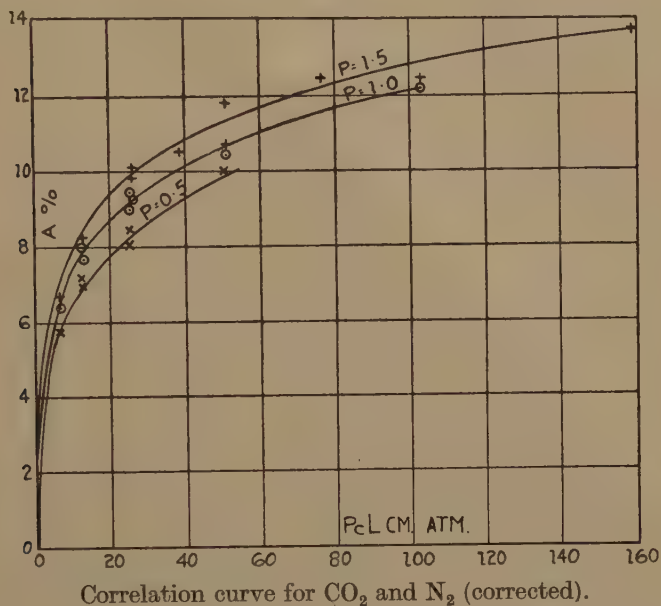
are in excellent agreement with those of the writer, as shown in Table III.

It was found that Beer's law holds, provided the total pressure is kept constant; this result confirmed the conclusions of several earlier investigators. In addition the dependence of the absorption on the total

pressure has been measured, and in fig. 7 the percentage absorption is plotted against the product, $p_c L$, for various values of the total pressure.

In absorption experiments the temperatures of the gas and source of radiation can be varied independently. If the temperature of the gas is increased at constant pressure, the number of absorbing molecules is reduced. Nevertheless, Hottel⁽¹¹⁾ found that increasing the temperature of carbon dioxide increased the percentage absorption for a fixed source of radiation. The increase of temperature raises some of the gas molecules to higher energy levels, and so broadens the bands by increasing the number of lines in them; the new levels may also have different intrinsic absorbing powers. These changes evidently increase

Fig. 7.



the absorption more than it is decreased by the reduction in the number of molecules.

Hottel⁽¹¹⁾ found the expression of his results simpler if compared for the same density, *i. e.* for the same number of absorbing molecules. The suggested formula was

$$A_c \propto \theta^{0.65}, \quad \dots \dots \dots (7)$$

where θ is the absolute temperature. In the present work the temperature of the carbon-dioxide was varied from 27° C. to 165° C., and within this range the results agreed reasonably well with (7).

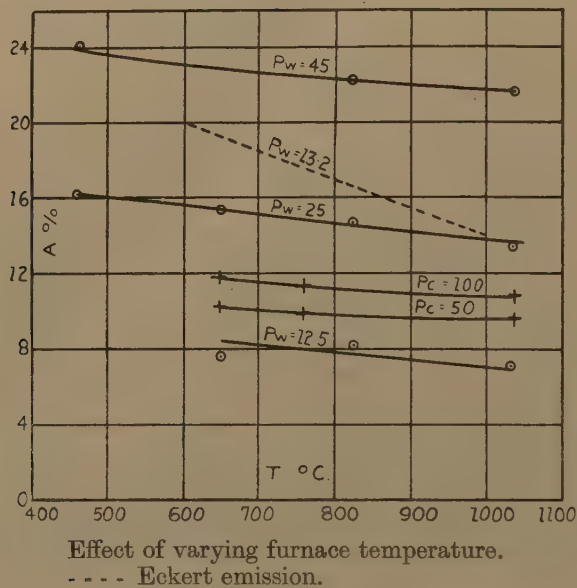
Changing the temperature of the source of radiation influences the absorption by altering the fraction of the radiation falling within the absorption bands. In the temperature range investigated, there is a

decrease of the percentage absorption with an increase of the temperature of the source, as shown in fig. 8.

In experiments on the emission of radiation by gases, the temperature of the gas and source are identical, and cannot be varied independently. Kirchhoff's law shows that the change in emissivity caused by changing the temperature from one value T' to another T'' is the same as the sum of the changes in the percentage absorption produced by changing (a) the temperature of the source, and (b) the temperature of the gas from T' to T'' .

Applying this result to the measured values for carbon dioxide, we have the experimental facts that there is an increase in the percentage absorption, when the temperature of the gas is increased, but a decrease

Fig. 8.



in the percentage absorption when the temperature of the source is increased (see fig. 8). These two variations are numerically of about the same order, and have opposite signs, so that the emissivity of carbon dioxide should be almost independent of temperature. The emissivity of carbon-dioxide has been measured at various temperatures by both Hottel⁽¹¹⁾ and Eckert⁽¹³⁾, who found that it was almost constant. Thus the above theory appears to be justified.

(b) *Water-Vapour and Mixtures of Water-Vapour and Nitrogen.*

The measurements of the absorption of radiation by water-vapour and mixtures of water-vapour and nitrogen are chiefly for conditions not previously investigated. Hottel made a few such measurements of

absorption, but since he used different black-body temperatures from the writer, the results cannot be easily compared.

The results for water-vapour and mixtures of water-vapour and nitrogen are complicated by the fact that Beer's law breaks down not only when the total pressure is varied, but even when the total pressure is constant, although the departures from Beer's law are smaller in the latter case. Table IV. gives some idea of the magnitudes of these departures. It can be seen that the departures from Beer's law when the total pressure is constant, are only about a quarter of those found by Eckert in the case of emission. Since the present experiments were completed, a paper by Hottel and Egbert⁽²⁰⁾ has appeared. In this paper they criticise Eckert for assuming that large departures from Beer's law occur at high temperatures of the order 1000° C., when the existence of these departures had been substantiated only at temperatures below 400° C. Considering all the available evidence, including some new measurements of their own on the emissivity of a dilute mixture

TABLE IV.

Percentage absorption for mixtures of water-vapour and nitrogen, showing typical departures from Beer's law.

L. cm.	p_w . cm. Hg.	p_n . cm. Hg.	A_w . Per cent.	p_w . cm. Hg.	p_n . cm. Hg.	A_w . Per cent.
25.4	76.0	—	26.6	—	—	—
50.8	38.0	—	24.6	38.0	38.0	26.4
101.6	19.0	—	21.9	19.0	57.0	24.6

of water-vapour and nitrogen in a tube 14 ft. long, they conclude that at high temperatures the departures from Beer's law are much smaller than supposed by Eckert. They suggest a new method of calculating the corrections in any particular case, and the values of these corrections agree well with those found by the author for the percentage absorption.

While the present investigation shows that the departures from Beer's law are small, when the source is hot (1000° C.) and the mixture of water-vapour and nitrogen comparatively cool (100° to 150° C.), the application of Kirchhoff's law to Eckert's results shows that if the source is also cool the deviations from Beer's law should be large. If Hottel's arguments are accepted, a similar use of Kirchhoff's law indicates that if the temperatures of the source and of the gas mixture responsible for the absorption are both high, the departures from Beer's law are again small. Thus the departures from Beer's law are practically unaffected by a change of the gas temperature from 100 to 1000° C., whereas a similar change in the temperature of the source reduces the magnitude of the departures by about 75 per cent. Such an increase in the temperature of the source increases the proportion of the radiation absorbed by the short-wave

water-band ($2.5-3.3\mu$) relatively to that absorbed by the longer wave-band ($5.0-8.5\mu$). The variation in the departures from Beer's law should therefore be sought in some difference in the structures of these bands. The difference could take the form of a difference in the half-widths of the lines in the two bands, but differences in the intensities and spacings of the lines would probably be sufficient to account for the variation in the departures from Beer's law. Present knowledge of the band-structures and line-widths is scarcely sufficient to justify an attempt to say which of these explanations is correct.

As shown in fig. 8, the percentage absorption by water-vapour of radiation from a black-body source, decreases slowly as the temperature of the source is increased from 500° to 1000°C . This decrease is, however, only about half as rapid as the decrease in emissivity of water-vapour between these temperatures. Using Kirchhoff's law in the same way as for carbon dioxide, the variation of the percentage absorption with the temperature of the water-vapour can be estimated. As the absorption experiments available for this calculation used water-vapour alone, whereas the comparable emission experiments used mixtures of water-vapour and nitrogen, the estimate may not be perfectly accurate, but it should give at least approximate results. The conclusion is that for a constant source of radiation, the percentage absorption due to water-vapour at constant pressure should decrease with decreasing gas temperature. It should be added that the experimental evidence from which this deduction was made used only values of p_w less than one atmosphere. Now according to Hottel the percentage absorption of water-vapour at constant pressure is independent of the gas temperature, although he admitted that his absorption experiments were not very accurate.

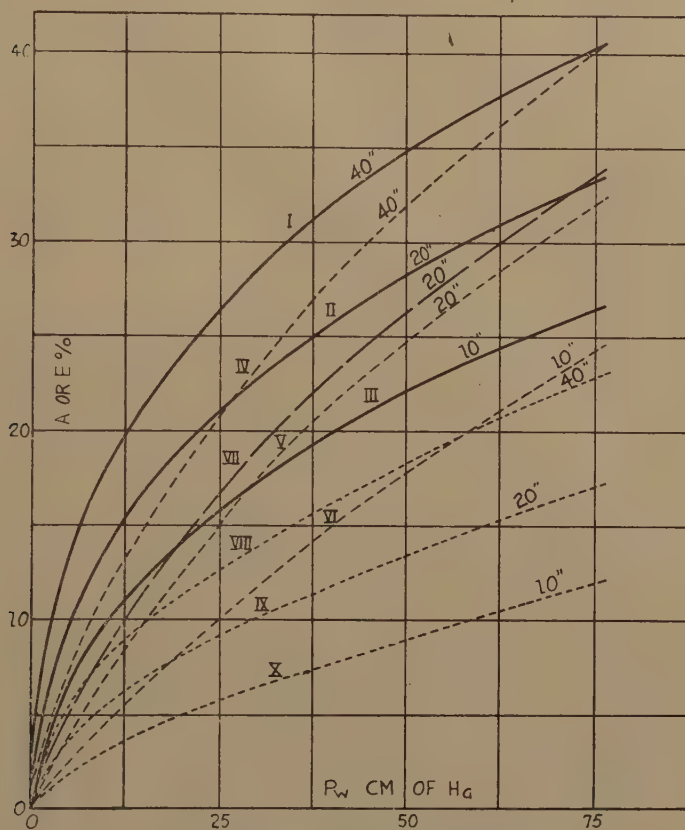
It would have been interesting to have measured the percentage absorption by water-vapour at widely different temperatures, but in the present investigation the temperature was not reduced below 110°C ., for fear of condensation, or raised above 165°C ., because the apparatus contained rubber. This temperature range was insufficient to form any conclusion on the change of absorption with gas temperature. Nevertheless the values of the percentage absorption obtained can be compared with Hottel's values of the emissivity at the temperature of the source of radiation in the absorption experiments. By Kirchhoff's law it follows that the difference between these values of the percentage absorption and emissivity is equal to the change in the percentage absorption, when the temperature of the water-vapour is raised to the temperature of the source of radiation. Fig. 9 contains corresponding values of the emissivity according to Hottel, and also according to Eckert, and of the percentage absorption according to the author. From these values it appears that the percentage absorption does not vary with the temperature of the water vapour when $p_w=1$ atmosphere, but decreases with increasing gas temperature for smaller values of the partial pressure of water-vapour.

Fishenden⁽²¹⁾ found that the absorption due to a gas mixture con-

taining 7.5 per cent. carbon-dioxide and 20.5 per cent. water-vapour was independent of its temperature between 200° and 500° C. As the absorption due to carbon dioxide increases with temperature, this result is further evidence for a decrease of the percentage absorption due to water-vapour with increasing gas temperature.

This question evidently requires further direct investigation, but it would probably be better to compare the absorption at different temperatures for the same density rather than for the same pressure; Hottel

Fig. 9.



Comparison of various investigations.

$p_w + p_n = \text{one atmosphere.}$

I., II., III. McCaig absorption. IV., V., VI. Eckert emission.

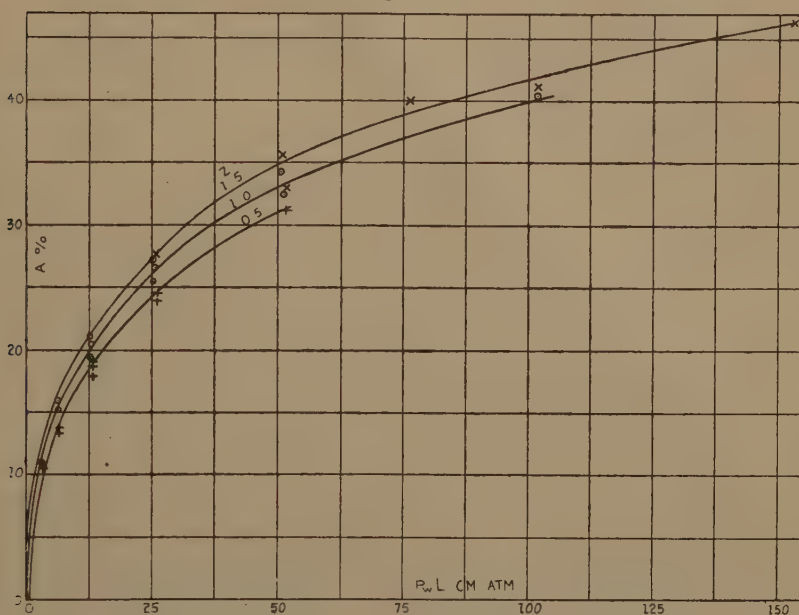
VII. Hottel emission. VIII., IX., X. McCaig absorption, but some density as in emission experiments.

found this method more convenient in the case of carbon dioxide. Obtaining the absorption for the higher gas temperature by applying Kirchhoff's law to Hottel's figures, and that for the lower temperature directly from the present investigation, it appears from fig. 9 that for any constant density of water-vapour, the percentage absorption is

In view of the uncertainty concerning the true values of a and b , it appears preferable to find the values that give the best correlation. Taking a as 0.65 (p_w and p_n being measured in atmospheres) there is little to choose between the correlation curves drawn with $b=0.4$ and 0.2. The higher value, which is the one theoretically expected, gives the better correlation for low nitrogen pressures, and the smaller value leads to better results for high nitrogen pressures. Finally, to simplify the use of the curves, δ is replaced by an arbitrary parameter z , chosen so that a has the value unity, while b is taken as 0.5. Thus the absorption for any given mixture of water-vapour and nitrogen can be read from fig. 8, after calculating $p_w L$ and z , where z is given by the formula

$$z = p_w + 0.5p_n, \quad \dots \dots \dots (9)$$

Fig. 10.



Correlation curves for water-vapour and nitrogen.

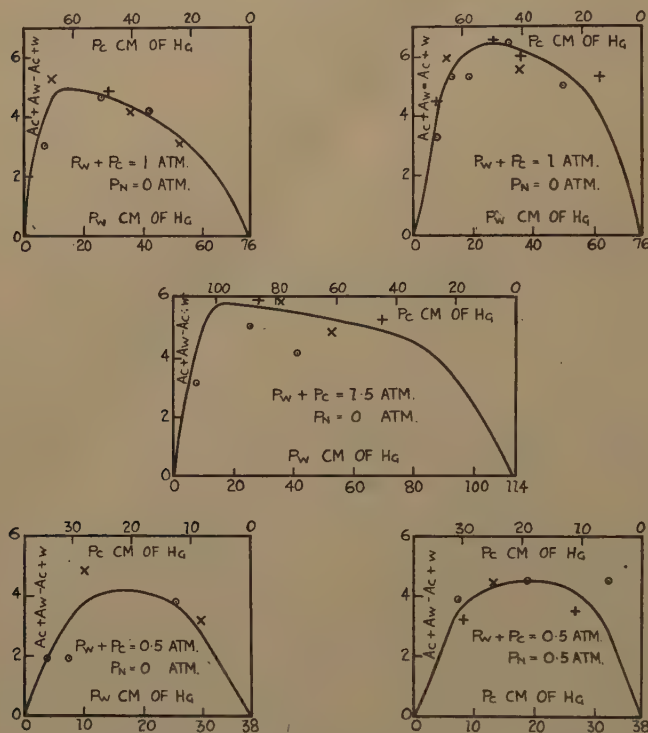
where p_w and p_n are in atmospheres. These correlation curves are shown in fig. 10. This figure gives the results for mixtures of water-vapour and nitrogen with reasonable accuracy, while compressing them into a small compact family of curves.

(c) *Mixtures of Water-Vapour and Carbon-Dioxide.*

When water-vapour and carbon-dioxide are both present, each gas increases the absorption of the other by pressure broadening. On the other hand, the absorption bands of the two gases overlap to some extent, causing a reduction in the absorption due to the mixture. These two opposing effects cannot be studied separately from each other. All

that can be attempted at present is to measure the absorption for various gas mixtures, and suggest some empirical method of deriving the absorption due to a given mixture from the sum of the absorptions of the constituents. In fig. 11, $A_c + A_w - A_{c+w}$ is plotted against the partial pressures p_c and p_w . The value of this difference involves the experimental and interpolation errors of three quantities much larger than itself, and is thus not very accurate. The correction increases slightly with the total pressure, but rather surprisingly appears to be about the same for all three tubes. Thus it appears that for a given

Fig. 11.



Corrections for water-vapour and carbon dioxide.

0=40 in. tube ; +=20 in. tube ; x=10 in. tube.

total pressure, all the overlapping that can occur is attained with the shortest tube. An increase in line-width consequent on an increase of pressure does, however, increase the overlapping.

Hottel, who has investigated the emission of radiation by mixtures of water-vapour and carbon-dioxide at a total pressure of one atmosphere, expresses his correction as a percentage of $A_c + A_w$. He found the same obstacles to accuracy as did the author. In general the values of his corrections are less than half those found by the author; He used only one value of L , so his results give no evidence of the way in which the

correction depends on the path-length. The position with regard to the correction for band overlapping remains rather unsatisfactory, but fortunately even a 100 per cent. error in the correction makes only a small difference to the final result for the absorption of the mixture.

SUMMARY.

The absorption of black-body radiation by water-vapour, carbon dioxide, and mixtures of these gases with each other and with nitrogen has been measured. The total pressure was varied from 0.1 to 2.0 atmospheres, and the path-length from 25 to 100 cm. The primary object was to investigate the validity of Beer's law, which states that the absorption is a function of the product of the partial pressure of the absorbing gas and the path-length.

It was found that for mixtures of carbon-dioxide and nitrogen Beer's law holds provided the total pressure is constant, but that if the total pressure is increased, whether by compressing a fixed quantity of carbon-dioxide or by adding nitrogen, the law breaks down and the absorption increases. A similar increase was found with mixtures of water-vapour and nitrogen.

For water-vapour, even when the total pressure was kept constant, Beer's law failed when the partial pressure was increased, there being a slight increase in the absorption, but not nearly as great an increase as claimed by Eckert for the emitted radiation. Recently Hottel and Egbert have disputed Eckert's results, and it now appears that the considerable departures observed by Eckert occur only when the temperature of the gas for emission or the source of radiation for absorption is less than about 400° C.

The effect upon the absorption of changing the temperatures of the gas and the source of radiation, and the theoretical reasons for the departures from Beer's law, are discussed. Schnaidt's method of calculating the absorption has been found unsatisfactory for the present problems, but theoretical considerations have suggested a method of expressing the experimental results in a form convenient for practical purposes.

Bibliography.

- (1) F. Paschen, *Wied. Ann.* liii. p. 34 (1894).
- (2) K. Ångström, *Öfvers. Kongl. Vet. Akad.* Stockholm, lviii. p. 381 (1901).
- (3) K. Ångström, *Ann. der. Physik*, vi. p. 163 (1901).
- (4) Eva v. Bahr, *Verh. dtsch. physik. Ges.* xv. pp. 173, 710 and 1150 (1913).
- (5) J. Koch, *Öfvers Kongl. Vet. Akad.* lviii. p. 475 (1901).
- (6) C. Shaefer, *Ann. der Physik*, xv. p. 93 (1905).
- (7) G. Hettner, *Ann. der Physik*, lv. p. 476 (1918).
- (8) W. Lasareff, *Zeits. für Physik*, lxiv. p. 598 (1930).
- (9) A. Schack, *Zeits. für tech. Physik*, v. p. 267 (1924).
- (10) E. Schmidt, *Forschung. Ing. Wes. A*, iii. p. 57 (1932).
- (11) C. H. Hottel and H. C. Mangelsdorf, *Amer. Inst. Chem. Eng.* xxxi. p. 517 (1935).

- (12) E. Schmidt and E. Eckert, *Forschung. Ing. Wes.* A, lxxxvii. (1937).
- (13) E. Eckert, *Forschung. Ing. Wes.* B, cccxxxxvii. (1937).
- (14) H. Margenau and W. W. Watson, *Rev. Mod. Physics*, viii. p. 22 (1936).
- (15) W. W. Elsasser, *Astrophys. J.* lxxxvii. p. 497 (1938).
- (16) P. Schulz, *Phys. Zeits.* xxxix. p. 412 (1938).
- (17) S. D. Cornell and W. W. Watson, *Phys. Rev.* l. p. 279 (1936).
- (18) S. D. Cornell, *Phys. Rev.* li. p. 739 (1937).
- (19) W. Gerlach, *Ann. der Physik*, l. p. 233 (1916).
- (20) C. H. Hottel and R. B. Egbert, *Trans. Amer. Soc. Mech. Eng.* lxiii. p. 297 (1941).
- (21) M. Fishenden, *Engineering*, cxlii. p. 684 (1936).
- (22) F. Schnajdt, *Gerlands. Beiträge zur Geophysik*, liv. p. 203 (1939). (This paper contains a good bibliography up to 1938.)
- (23) G. Falckenberg, *Meteor. Zeits.* liii. p. 172 (1936).
- (24) G. Falckenberg, *Meteor. Zeits.* lv. p. 174 (1938).
- (25) H. H. Nielsen, *Phys. Rev.* lix. p. 565 (1941).

XXXVII. *The Reflexion of Electromagnetic Waves from a Parabolic Ionized Layer.*

By OLOF E. H. RYDBECK, Chalmers Institute of Technology,
Gothenburg, Sweden*.

[Received April 20, 1942.]

THE electron density distribution of the undisturbed F_2 -layer of the ionosphere very often is essentially parabolic^{(1), (2)}, at least at sufficient distances from the equator. It is therefore of general interest to study briefly the reflexion of electromagnetic waves from such a layer.

We assume that plane waves are transmitted in a direction normal to the ionized layer, the electron density of which is a function of the vertical (normal) distance only. The wave equation then gets the following form, viz.,

$$\frac{d^2\Pi}{dz^2} + \frac{\omega^2}{c^2} \cdot \epsilon \cdot \Pi = 0, \quad (1)$$

so that the electric field-strength (transversal) will be represented by

$$E = \Pi \cdot e^{-j\omega t} \cdot (j = \sqrt{-1}). \quad (2)$$

If the electron density distribution is parabolic with a half-thickness Δh_m , then

$$N = N_{\max} \left(1 - \left(\frac{z}{\Delta h_m} \right)^2 \right), \quad (2a)$$

* Communicated by the Author.

where N_{\max} is the maximum electron density and z , the vertical distance, is counted from its level (positive downwards). Remembering that the dielectric "constant" is *

$$\epsilon = 1 - \frac{4\pi N e^2}{m\omega^2} = 1 - \frac{f_{cm}^2}{f^2} \left\{ 1 - \left(\frac{z}{\Delta h_m} \right)^2 \right\}, \quad (\omega = 2\pi f) \quad (3)$$

with f_{cm} denoting the so-called critical or penetration frequency of the layer, it is easily established that the wave equation can be written

$$\frac{d^2\Pi}{dv^2} + \left[j\rho - \frac{v^2}{4} \right] \cdot \Pi = 0; \quad (4)$$

where

$$\rho = \pi \cdot \frac{\Delta h_m}{\lambda_{cm}} \cdot \frac{f_{cm}^2 - f^2}{f_{cm}^2} \quad \text{and} \quad v = \left(\frac{4\pi \cdot \Delta h_m}{\lambda_{cm}} \right)^{1/2} \cdot \frac{z}{\Delta h_m} \cdot e^{j\frac{\pi}{4}} = u \cdot e^{j\frac{\pi}{4}}. \quad (4a)$$

$\Delta h_m/\lambda_{cm}$ always is great, 2 to $5 \cdot 10^3$ is a typical day-time value for the F_2 -layer. ρ is also great except near the penetration frequency. Eq. (4) is satisfied by Weber's parabolic cylinder functions. One of them is †

$$\Pi = D_{j\rho - \frac{1}{2}} \left(u \cdot e^{j\frac{\pi}{4}} \right). \quad (5)$$

From Whittaker's integral representation one easily gets the following asymptotic expansion, viz.:

$$D_{j\rho - \frac{1}{2}} \left(u \cdot e^{j\frac{\pi}{4}} \right) \sim e^{-j\left(\frac{u^2}{4} - \rho \ln u + \frac{\pi}{8}\right)} \cdot \frac{e^{-\frac{\pi \cdot \rho}{4}}}{(u)^{\frac{1}{2}}} \cdot \left[1 - \frac{j(\rho + j\frac{1}{2})(\rho + j\frac{3}{2})}{2u^2} + \dots \right] \quad (5a)$$

As u always is great the only restriction is that ρ must be sufficiently small. A safe limit is marked by

$$\frac{\rho^2}{2u^2} \ll 1 \quad \text{or} \quad |\Delta f| = |f_{cm} - f| < \left(\frac{2c}{\pi \cdot \Delta h_m} \right)^{1/2} \cdot (f_{cm})^{1/2}.$$

If $f_{cm} = 10^7$, for example, and $\Delta h_m = 100$ km., then $|\Delta f|$ becomes less than about 1, $4 \cdot 10^5$ c/s. The present expansion, therefore, can only be used in the neighbourhood of the penetration frequency, f_{cm} . As this is the region where the main deviation from geometrical optics occurs, the expansion is sufficient for our present purpose.

Multiplying by the time-factor we see that

$$e^{-j\omega t} \cdot D_{j\rho - \frac{1}{2}} \left(u \cdot e^{j\frac{\pi}{4}} \right)$$

* m = mass of electron, e = charge of electron.

† The symbol is that used by Whittaker. See Whittaker & Watson 'A Course of Modern Analysis,' p. 347.

yields an up-going wave. Another solution is the conjugate wave function

$$D_{-j\rho-\frac{1}{2}}\left(u \cdot e^{-j\frac{\pi}{4}}\right) \sim e^{j\left(\frac{u^2}{4}-\rho \ln u+\frac{\pi}{8}\right)} \cdot \frac{e^{-\frac{\pi \cdot \rho}{2}}}{(u)^{\frac{1}{2}}} \cdot \overset{\text{phase angle } \phi}{\left[1+\frac{j(\rho-j\frac{1}{2})(\rho-j\frac{3}{2})}{2u^2}-\dots\right]}. \quad (6)$$

which multiplied by the same time-factor yields a down-coming wave. Next we have to find a suitable "circuit" relation of the differential equation connecting the incident, the reflected and the refracted wave. Fortunately enough, there exists a fairly simple one, viz.,

$$\underbrace{D_{j\rho-\frac{1}{2}}\left(u \cdot e^{-j\frac{\pi}{4}}\right)}_{\text{incident wave}} = \underbrace{\frac{\Gamma(j\rho+\frac{1}{2})}{(2\pi)^{\frac{1}{2}}} \cdot e^{\frac{\pi \cdot \rho}{2}+j\frac{\pi}{4}} \cdot D_{-j\rho-\frac{1}{2}}\left(u \cdot e^{-j\frac{\pi}{4}}\right)}_{\text{reflected wave}} + \underbrace{\frac{\Gamma(j\rho+\frac{1}{2})}{(2\pi)^{\frac{1}{2}}} \cdot e^{-\frac{\pi \cdot \rho}{2}-j\frac{\pi}{4}} \cdot D_{-j\rho-\frac{1}{2}}\left(u \cdot e^{j\frac{3}{4}\pi}\right)}_{\text{refracted wave}}. \quad (7)$$

Mathematically speaking, the above expression is the analytical continuation of the function of the refracted ray, represented by an asymptotic series for instance, beyond the region for which this special representation is valid. The law of refraction gives us, therefore, a physical visualization of the concept of analytical continuation.

From a practical point of view we are primarily interested in the phase difference between the up-going and down-coming waves at the bottom of the layer ($z=\Delta h_m$). This is

$$\Delta S = \frac{u^2}{2} - \rho \cdot \ln(4u^2) + \text{phase} \left[\frac{\Gamma(2j\rho)}{\Gamma(j\rho)} \right] + 2 \cdot \phi + \frac{\pi}{2}.$$

The time of travel of the returned ray times the velocity of light in vacuum generally is called the virtual path-length. This is equal to twice the virtual height (vertical incidence). The virtual height therefore becomes

$$\Delta h_v = \frac{c}{2} \cdot \frac{d(\Delta S)}{d\omega} = \frac{c}{4\pi} \cdot \left\{ -\ln(4u^2) + \frac{2\rho}{u^2} + \text{Re}[2\psi(2j\rho) - \psi(j\rho)] \right\} \cdot \frac{d\rho}{df},$$

where $\psi(a)$ is the logarithmic derivative of $\Gamma(a)$. This finally yields

$$\Delta h_v = \frac{\Delta h_m}{2} \cdot \frac{f}{f_{cm}} \cdot \left\{ \gamma + \ln \left(16\pi \frac{\Delta h_m}{\lambda_{cm}} \right) - \rho \left(\frac{2}{u^2} - \text{Re} \sum_{n=1}^{\infty} \frac{2\rho - j3n}{n(n+j\rho)(n+2j\rho)} \right) \right\}, \quad (8)$$

with γ denoting Euler's constant 0.5772. At the penetration frequency the exact virtual height becomes

$$\Delta h_v = \frac{\Delta h_m}{2} \cdot \left\{ \gamma + \ln \left(16\pi \frac{\Delta h_m}{\lambda_{cm}} \right) \right\}. \quad (8a)$$

This is not infinite as should follow from the classical virtual height expression of the geometrical optics, viz.,

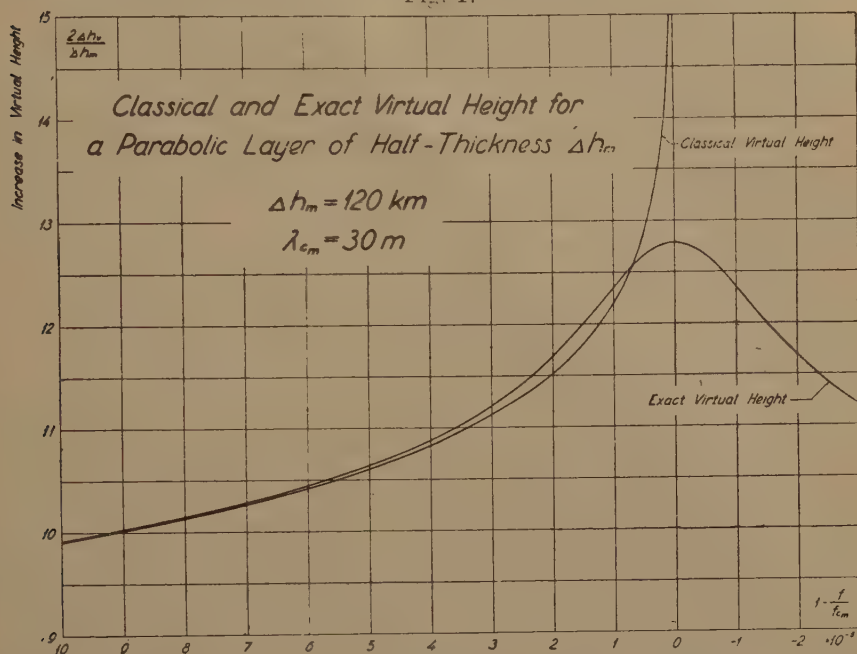
$$\Delta h_v = \frac{\Delta h_m}{2} \cdot \frac{f}{f_{cm}} \cdot \ln \left(\frac{f_{cm} + f}{f_{cm} - f} \right). \quad (8b)$$

It is fairly easy to show from Eq. (8) that

$$\lim_{\left(\frac{\Delta h_m}{\lambda_{cm}} \rightarrow \infty \right)} \frac{(\Delta h_v)_{\text{exact}}}{(\Delta h_v)_{\text{classical}}} = 1.$$

This shows mathematically the transition to diffraction-free optics when the geometrical dimensions become infinitely great. Fig. 1 shows a plot

Fig. 1.



of the classical and true virtual heights near the penetration frequency for a parabolic layer of 120 km. half-thickness and a critical wave-length of 30 m. For such a thick layer the discrepancy between the classical and exact virtual heights is appreciable only within a very narrow frequency range. For the layer in question this region is only about 1000 c/s wide. For a thinner layer this region becomes proportionately wider. For a very thin layer with a half-thickness of only four wave-lengths, i. e., 120 m., the discrepancy region becomes considerable, viz. 10^6 c/s. For such a thin layer the reflexion coefficient will also differ considerably from the classical one.

A few words should be said about the introduction of the electronic collisional frequency, ν . If $\nu^2 \ll \omega^2$, it is formally introduced in the wave function if f_{cm} is given the phase angle $-\alpha$, where $\alpha = \frac{1}{2} \arctan \left(\frac{\nu}{\omega} \right)$.

The introduction of the collisional frequency actually complicates the numerical calculation very much. A treatment of this case therefore is outside the scope of the present communication*. It should suffice here to show the numerical results for a fairly thick and a thin layer. By means of (4) and (4a) the corresponding results for layers of other

Fig. 2.

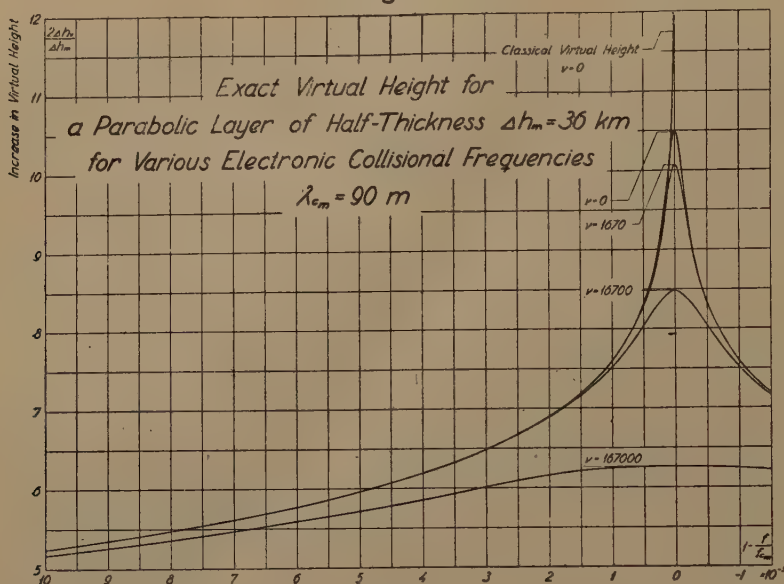
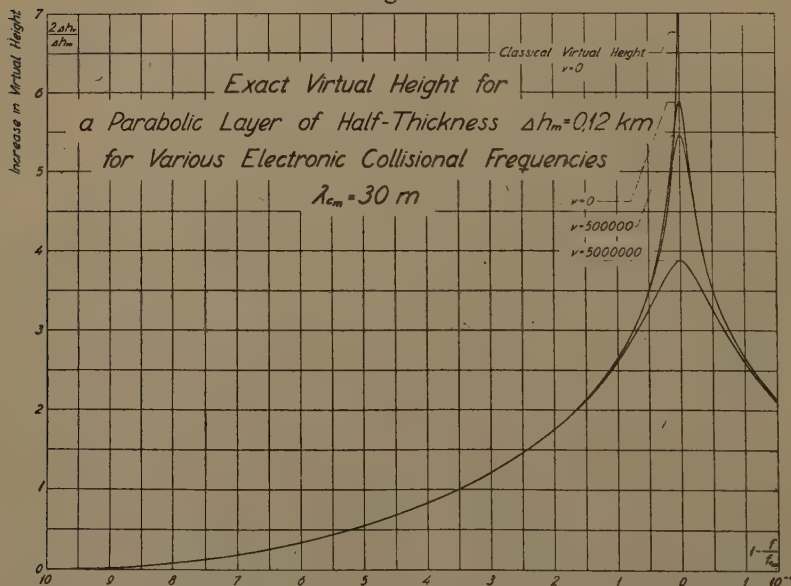


Fig. 3.

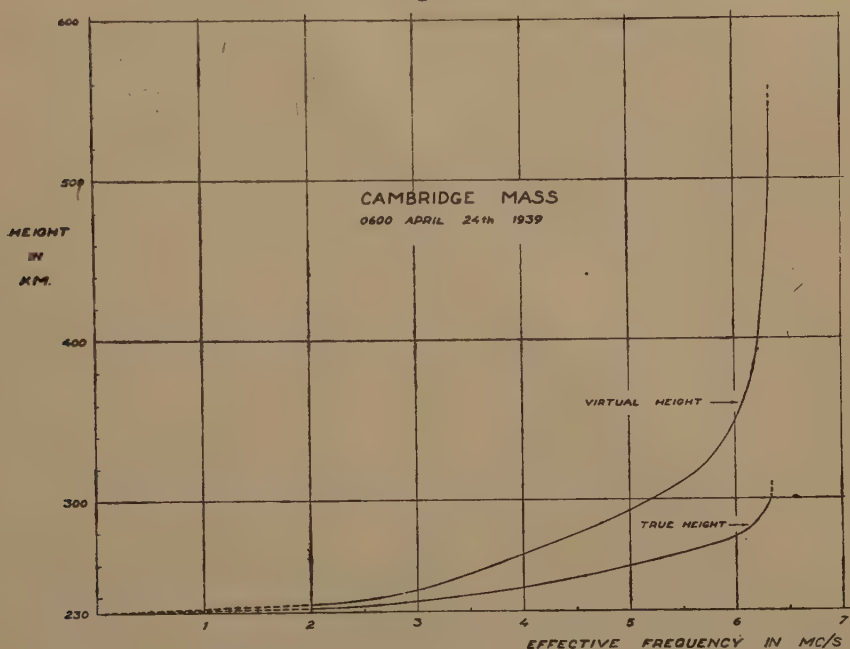


* A complete treatment of the dissipative case will appear shortly in the Transactions of Chalmers University.

dimensions can be obtained by transformation from the plots on fig. 2 and fig. 3. It is shown by fig. 3 that a considerable collisional frequency is necessary to cause any appreciable reduction in virtual height for the thin layer.

Fig. 4 shows a low-loss sensitive recording of ionospheric reflexions at Cambridge, Mass., from a practically parabolic layer. The layer half-thickness is about 70 km. and the critical wave-length is 47.2 m. The exact no-loss virtual height at the critical frequency becomes $5.88 \Delta h_m$. The highest virtual height recorded is $4.60 \Delta h_m$. If we transform the results on fig. 2 to the case $\Delta_m = 72$ km. and $\lambda_{cm} = 45$ m., we find that the

Fig. 4.



measurements just referred to indicate a collisional frequency less than that about 9000.

When the medium is non-dissipative one immediately obtains from the circuit relation that the reflexion coefficient, R , is determined by

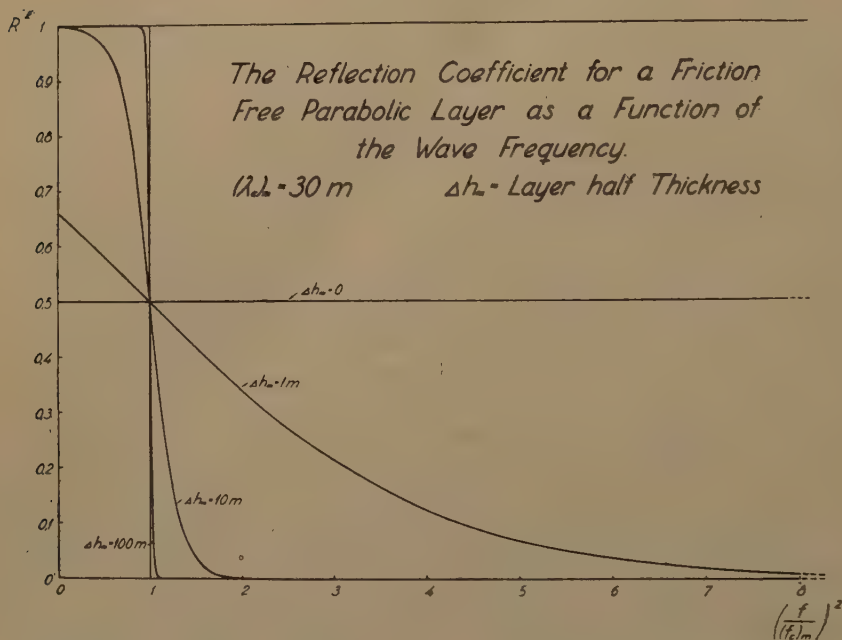
$$\frac{R^2}{1-R^2} = e^{2\pi p} = e^{4\pi^2 \cdot \frac{\Delta h_m}{\lambda_{cm}} \cdot \frac{f_{cm}^2 - f^2}{2f_{cm}^2}} \dots \dots \dots (9)$$

This holds throughout the frequency range. Fig. 5 shows several plots of R^2 as a function of $(f/f_{cm})^2$. The critical wave-length is 30 m. as before. The deviation from classical optics is practically noticeable first for a half-thickness of about four to three wave-lengths. As the layer becomes even thinner appreciable reflexions appear at frequencies well above the critical frequency. It is obvious that the critical frequency conception is misleading for a very thin layer.

Finally it should be stated that results similar to Eq. (8) have been obtained by Rawer⁽³⁾ in an excellent paper on the reflection of electromagnetic waves from dissipative Epstein-layers. A result closely similar to (9) for the penetration frequency region has been communicated by Booker to Appleton⁽⁴⁾.

An extended study of the parabolic solutions for other values of frequency necessitates the expansion of the wave functions in asymptotic series which can be used when not only u but also ρ is great. Interesting as such a study may be, it is outside the scope of the present communication. This is especially true because the discrepancy between the classical and exact theories is entirely negligible for lower frequencies, as is clearly

Fig. 5.



demonstrated by figs. 1, 2 and 3. An account of these expansions will, however, appear shortly in the Transactions of Chalmers University, Gothenburg, Sweden, to which those especially interested are referred.

References.

- (1) O. Rydbeck, "The Propagation of Electromagnetic Waves in an Ionized Medium and the Calculation of the True Heights of the Ionized Layers of the Atmosphere." *Phil. Mag.* ser. 7, vol. xxx. p. 282 (Oct. 1940).
- (2) O. Rydbeck, "A Theoretical Survey of the Possibilities of Determining the Distribution of the Free Electrons in the Upper Atmosphere." *Transactions of Chalmers University*, no. 3, Gothenburg, Sweden (1942).
- (3) K. Rawer, "Elektrische Wellen in einem geschichteten Medium." *Ann. d. Phys.* xxxv. p. 402 (1939).
- (4) E. V. Appleton, R. Naismith and L. J. Ingram, "The Critical Frequency Method of Measuring Upper-Atmospheric Ionization." *Proc. Phys. Soc.* li. p. 81 (1939).

XXXVIII. *The Single Potential Difference at a Cadmium Electrode.*

By J. ALAN CHALMERS, M.A., Ph.D.

Lecturer in Physics in the Durham Colleges of the University of Durham *.

[Received January 18, 1943.]

1. *Introduction.*

"SINGLE potential differences" at electrodes of metals in electrolytes have often been discussed and many attempts have been made to measure these. In earlier work it was thought that single potential differences could be identified with the "electrode potential" of electrochemistry, if these could be given absolute values; if the "absolute" potential difference at any one electrode could be found, it was thought that the potential difference at any other electrode could be found by using the measured potential difference of a cell containing these two electrodes. Attention was focused on the determination of the "absolute zero" of potential, and divergent results were obtained.

More recently it has been realised that the metal-metal contact cannot be ignored; the author ⁽¹⁾⁻⁽⁷⁾ in a series of papers has discussed contact potentials and has pointed out that the single potential difference at an electrode is what is termed an "external" potential difference, and, since external potential differences exist at metal-metal contacts, these must be included in adding potential differences for the whole cell.

Klein and Lange ⁽⁸⁾ have used the result that the metal-metal potential difference is equal to the difference of the photoelectric work functions for the two metals, and so have obtained the extra term needed to deduce single potential differences from one known single potential difference and measured potential differences of cells. It is necessary only to measure one single potential difference to obtain others by this method, if the photoelectric work functions are known with sufficient accuracy.

The methods of obtaining the single potential difference at an electrode can be divided into two main classes, which can be described as "electrokinetic" and "electrostatic." In the electrokinetic methods, attention is paid to the potential difference across the actual boundary of relative motion between electrode and electrolyte. In many of these methods conditions are arranged so that there is no potential difference at one electrode, and the completion of the cell gives a measured potential difference from which is obtained the potential difference at the other electrode, using, where necessary, the metal-metal potential difference.

* Communicated by the Author.

An electrode at which there is no potential difference is called a "null" electrode. In using electrokinetic methods, it is assumed that the only boundary at an electrode at which there is a potential difference is the boundary of relative motion. Discrepancies which have been found have been ascribed to the fact that there may be other boundaries with potential differences; in certain cases this must be so, for two "null" electrodes, with the same metal, do not give zero potential difference; but in other cases the discrepancies are due to neglect of the metal-metal potential difference.

The electrostatic method of determining the single potential difference consists of measuring the metal-electrolyte Volta effect and assuming that there is no potential difference at the metal-air and electrolyte-air boundaries. The author⁽⁵⁾ has discussed this matter and has pointed out that a potential difference exists at an electrolyte-air boundary (Chalmers and Pasquill⁽⁹⁾). Klein and Lange⁽⁸⁾ used this electrostatic method to obtain the single potential difference in one case, to be used with metal-metal potential differences and potential differences of cells to give other single potential differences. This method has the disadvantage of all Volta effect methods in that surface films on the metal-air surface have a considerable influence on the results obtained.

The results obtained by Klein and Lange⁽⁸⁾ were not in agreement with results obtained by electrokinetic methods and this would strengthen the argument that extra boundaries exist. But the author⁽⁶⁾ has shown that a consistent scheme can be drawn up to fit the results of electrokinetic measurements with mercury in acid or potassium salt solutions, mercury in mercurous salt solutions and silver in silver salt solutions as well as Klein and Lange's results by the electrostatic method, provided we assume: (1) the air-water potential difference found by Chalmers and Pasquill⁽⁹⁾, (2) an extra boundary, perhaps due to an "amalgam," at an electrode of mercury in acid or potassium salt solutions, and (3) a small change of the mercury-silver potential difference from that used by Klein and Lange⁽⁸⁾. The consistency achieved is a sufficient justification for these assumptions, but it would clearly be preferable if further experimental results could be included; further measurements of metal-electrolyte Volta effects could only verify assumption (3) above.

Electrokinetic measurements of any accuracy appear to have been made only with mercury and silver, discussed already, with liquid gallium and with amalgams. In the case of gallium, discussion is handicapped by the lack of accurate knowledge of the photoelectric work function for liquid gallium; in the case of amalgams there is the difficulty of distinguishing between the effects of the mercury and the other metal. It is only in the cases of liquid metals that it is possible to obtain results when a definite potential difference exists at an electrode; in other cases, electrokinetic methods so far devised are not applicable except when a null electrode is obtainable, and are therefore only available in the cases of the few metals for which a null solution of its salt lies between the concentration of saturation and a concentration of about 10^{-4} N or

10^{-5} N, beyond which the metal ions have little effect (McAulay and Spooner⁽¹⁰⁾). Silver answers these requirements, the null solution being between 10^{-4} N and 10^{-5} N, but it seems probable that few other metals do so. Using the results of Klein and Lange⁽⁸⁾, it would appear that the only other metals likely to satisfy the condition would be cadmium and thallium in thalious salts. A simple electrokinetic method was therefore tried for a cadmium electrode.

2. The Scraped Electrode.

Bennewitz and Schultz⁽¹¹⁾ and Bennewitz and Bigalke⁽¹²⁾ have used the scraping of an electrode to find the "null" solution of silver nitrate in contact with a silver electrode. If there is a double layer of charge at the electrode, then this will be removed mechanically by scraping the electrode; if a circuit is completed with a similar unscraped electrode and a galvanometer, then there will be a current through the galvanometer when the double layer is renewed and it is clear that this current will alter its sign when the signs of the double layer are reversed, and there will be no current on scraping when there is no double layer. It does not seem to be possible at all easily to apply a potential difference across the layer and so eliminate the natural potential difference as is done in the capillary electrometer, and therefore it is only possible to use the method when a null solution is obtainable.

The experimental arrangement for determining the single potential difference at a cadmium electrode was very simple. A stick of cadmium metal was cut in two and connected by copper wires to a galvanometer; the two ends were placed into a solution, care being taken to keep the cadmium-copper junctions dry, and then either piece of cadmium could be scraped by means of a piece of glass tube. It was found that a table galvanometer of the "Pye" type was sufficiently sensitive for the purpose. It was found that for all concentrations of cadmium sulphate equal to or less than normal there was a current of positive charge *from* the scraped electrode after scraping, so that the double layer at the electrode must have its negative side on the metal. For concentrations greater than twice normal, there was a current of positive charge *to* the scraped electrode, and so a double layer with positive side on the metal. For a concentration of $1\frac{1}{2}$ normal it is not possible to determine the direction of the current. Since a change from N to $1\frac{1}{2}$ N only alters the potential difference by 0.005 volts according to the Nernst theory, it is clearly not necessary to determine the concentration of the "null" solution more accurately than this.

To the degree of accuracy which is obtainable in measurements of this kind, we can say that a normal cadmium sulphate solution is a "null" solution, and any difference between normal strength and unit activity can also be neglected. The results quoted by Klein and Lange⁽⁸⁾ give the potential difference of a cell with the two electrodes Ag./N.AgNO_3 and Cd./N.CdSO_4 as 1.20 volts, the silver being positive. As already determined⁽⁶⁾, the electrode Ag./NAgNO_3 has a potential difference of

0.25 volts, so that there remains a potential difference of 0.95 volts at the metal-metal junction; the values of the photoelectric work functions of silver and cadmium used by Klein and Lange⁽⁸⁾ give a metal-metal potential difference of 1.00 volts, so that agreement is quite satisfactory. Since the author⁽⁶⁾ has used a mercury-silver potential difference of 0.10 volts instead of 0.16 given by Klein and Lange, an alteration of the photoelectric work function of silver from 4.68 volts to 4.62 volts would bring all the results into accord and the present results could be included in fig. 1 of the author's paper⁽⁶⁾ by a further column headed "normal cadmium sulphate" with "cadmium" on the zero line, "silver" at -0.95 and "mercury" at -0.85.

If the curves of fig. 6 in Klein and Lange's⁽⁸⁾ paper are moved to the right by about 0.33 volts, the cadmium results agree with what is to be expected. This difference of 0.33 volts is to be attributed to the air-water potential difference of 0.26 volts (Chalmers and Pasquill⁽⁹⁾), together with a correction from Klein and Lange's figures for argon to those for moist air as used in other measurements.

It may be noted that a cell with a null cadmium electrode and a calomel electrode gives a potential difference of 0.68 volts. If, therefore, the experiments with cadmium had been carried out before the effect of the metal-metal contact had been realised, a value would have been obtained for the single potential of the calomel electrode of 0.68 volts, to be compared with 0.56 volts from the capillary electrometer and other electrokinetic measurements with mercury in acid or potassium salt solutions, -0.15 volts from measurements with mercury in mercurous salt solutions, and -0.25 from the silver measurements.

Since the present results with cadmium are consistent with results for mercury and silver, there seems further justification for the assumptions mentioned above in regard to the air-water potential difference and the "amalgam." The results also show that, with cadmium as with silver, the electrokinetic potential difference is the same as the total potential difference across the electrode, and there are no other double layers present; this could be verified by measuring the Volta effect for cadmium and normal cadmium sulphate, which we would expect to be 0.25 volts in moist air or 0.33 volts in argon.

Summary.

By the method of the scraped electrode, it has been found that the "null" solution for a cadmium electrode in a solution of cadmium sulphate is of concentration about $1\frac{1}{2}$ N. Cadmium in a normal solution has a single electrode potential of about -0.005 volts, which can be taken as zero to the degree of accuracy of other results. This value is consistent with the others obtained by the author⁽⁶⁾.

References.

- (1) Chalmers, *Phil. Mag.* xxxiii. p. 399 (1942).
- (2) Chalmers, *Phil. Mag.* xxxiii. p. 416 (1942).
- (3) Chalmers, *Phil. Mag.* xxxiii. p. 496 (1942).

- (4) Chalmers, Phil. Mag. xxxiii. p. 506 (1942).
- (5) Chalmers, Phil. Mag. xxxiii. p. 594 (1942).
- (6) Chalmers, Phil. Mag. xxxiii. p. 599 (1942).
- (7) Chalmers, Phil. Mag. xxxiii. p. 608 (1942).
- (8) Klein and Lange, *Zeit. Elektrochem.* xliii. p. 570 (1937).
- (9) Chalmers and Pasquill, Phil. Mag. xxiii. p. 88 (1937).
- (10) McAulay and Spooner, P. R. S. cxxxviii. p. 494 (1932).
- (11) Bennewitz and Schulz, *Zeit. Phys. Chem.* cxxiv. p. 115 (1926).
- (12) Bennewitz and Bigalke, *Zeit. Phys. Chem.* cliv. p. 113 (1931).

XXXIX. *The Stage Efficiency of an Impulse Generator.*
A Mathematical Note.

By G. H. RAWCLIFFE (Robert Gordon's Technical College, Aberdeen,
formerly of Liverpool University), M.A., A.M.I.E.E.*

[Received November 17, 1942.]

Introduction.

THE prototype impulse generator, from which all impulse generators are derived, is a condenser C charged from an independent direct current source V and discharged through a resistance R across which the impulse voltage is developed. As the condenser is connected across the resistance, the voltage rises (theoretically) instantaneously to the condenser charging voltage and then decays exponentially to zero. The time taken for the voltage to fall to half the original can readily be shown to be $CR \log_e 2$ seconds. In practice the condenser and resistance are "connected" by the breakdown of a spark.

An impulse voltage wave is defined as a (t_1/t_2) wave, where t_1 is the time to the peak voltage and t_2 is the time to half voltage on the wave tail, both in microseconds. The prototype impulse wave would thus be a $(\text{Zero}/CR \log_e 2 \times 10^6)$ wave, according to this definition.

Effect of Inductance.

If inductance is introduced into the circuit (in fact, a circuit without any inductance is a mathematical abstraction) the effect is two-fold.

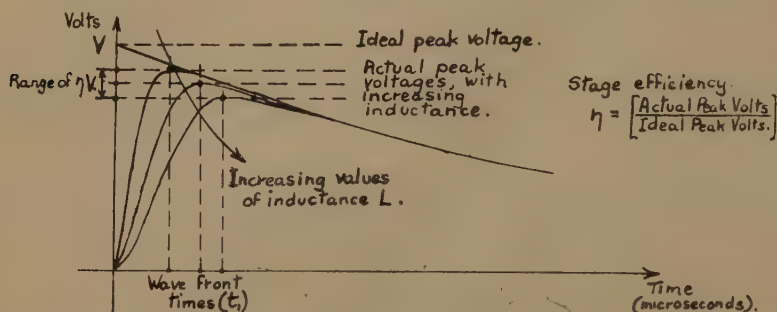
1. The peak voltage on the discharge resistance takes a finite time t_1 to develop, instead of appearing instantaneously.
2. The value of the peak voltage so reached is less than the condenser charging voltage. As the inductance rises the peak becomes smaller and the time to reach it greater.

A family of impulse voltages for increasing values of inductance is shown in fig. 1. The magnitude of both effects is exaggerated for clarity of illustration.

* Communicated by Professor E. W. Marchant, D.Sc.

It is desired to obtain simple mathematical expressions, susceptible of being rapidly evaluated for particular values of the constants concerned, for the time taken for the peak voltage to develop, and for the ratio between this peak voltage and the voltage to which the condenser was initially charged. The latter ratio is known as the stage efficiency of the impulse generator.

Fig. 1.



Sketch showing effect of increased values of inductance (L) on wave form.

Equation of Impulse Wave.

Considering a closed circuit of resistance, capacitance and inductance (fig. 2), as is effectively the case when the spark gap has broken down, we have the differential equation

$$L \frac{di}{dt} + Ri + \frac{q}{C} = 0$$

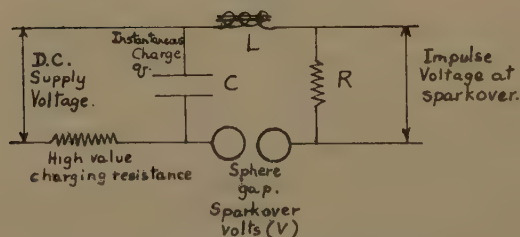
or

$$L \frac{d^2i}{dt^2} + R \frac{di}{dt} + \frac{i}{C} = 0,$$

since

$$i = \frac{dq}{dt}.$$

Fig. 2.



Sketch of prototype impulse generator.

We have also the conditions that at

$$t = 0,$$

$$i = 0,$$

and

$$\frac{q}{C} = -L \frac{di}{dt} = V = \text{the stage voltage}$$

$$= \text{the spark-gap breakdown voltage.}$$

Provided $R^2 \gg \frac{4L}{C}$, as is always the case since L is very small, we have the solution

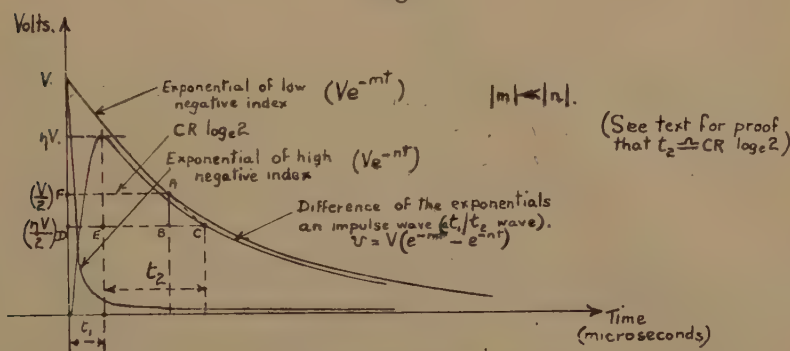
$$iR = \text{impulse voltage} = \frac{VR}{2L \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}} (e^{-mt} - e^{-nt})$$

$$= V \left(\frac{m+n}{n-m} \right) [e^{-mt} - e^{-nt}],$$

where m and $n = \frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$ respectively, and $|n| \gg |m|$.

This result will be seen on inspection to consist of the difference of two exponentially decaying quantities, of equal initial magnitude, but

Fig. 3.



Impulse wave shown as difference of two exponentials.

different exponential index. Fig. 3 shows that this difference does, in fact, represent graphically an impulse wave of the type it is desired to produce, rising rapidly to a peak value and decaying (relatively) slowly to zero. The physical and mathematical aspects of the problem are therefore seen accurately to correspond.

Time to the Peak of the Impulse Wave.

By simple differentiation and equation to zero the peak value of this impulse voltage can be shown to occur at time t_1 given by

$$t_1 = \left(\frac{\log_e \frac{n}{m}}{n-m} \right)$$

and hence the peak value of the voltage is given by substituting this value of time in the equation for the impulse voltage given above.

We have then that the impulse peak voltage is

$$V \left(\frac{m+n}{n-m} \right) \left[e^{-\left(\frac{m \log_e \frac{n}{m}}{n-m} \right)} - e^{-\left(\frac{n \log_e \frac{n}{m}}{n-m} \right)} \right],$$

the coefficient of V in the last expression being the stage efficiency η .

Approximations for Ready Computation.

The computation of either of these quantities t_1 and η for numerical values of LC and R is, however, laborious, particularly for η , and it is desired to reduce both these expressions to forms which permit quick computation. An approximation may be made as follows, based, as stated above, on the assumption, which is always justified for any real impulse generator, that $R^2 \gg \frac{4L}{C}$.

*Approximation for t_1 .**

We have

$$(n-m) = 2 \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \\ = \frac{R}{L} \left(1 - \frac{4L}{CR^2} \right)^{\frac{1}{2}}$$

expanding by the binomial theorem

$$\simeq \frac{R}{L} \left(1 - \frac{2L}{CR^2} \right), \\ \simeq \frac{R}{L}.$$

By similar binomial expansion, $\frac{n}{m}$ may be determined.

$$\frac{n}{m} \simeq \left(\frac{CR^2}{L} - 1 \right) \\ \simeq \left(\frac{CR^2}{L} \right)$$

thus we have

$$t_1 = \left(\frac{\log_e \frac{n}{m}}{n-m} \right) \simeq \left(\frac{L}{R} \log_e \frac{CR^2}{L} \right).$$

Approximation for η , the Stage Efficiency.

We have first to simplify

$$e^{-\left(\frac{m \log_e \frac{n}{m}}{n-m} \right)} \quad \text{and} \quad e^{-\left(\frac{n \log_e \frac{n}{m}}{n-m} \right)}.$$

* The author owes this paragraph, ultimately, to Professor W. J. John.

Now $\left(\frac{\log_e \frac{n}{m}}{n-m}\right)$ has already been shown to be approximately

$$\left(\frac{L}{R} \log_e \frac{CR^2}{L}\right),$$

and m and n can readily be seen, by binomial expansion, to be

$$\left(\frac{1}{CR}\right) \quad \text{and} \quad \left(\frac{R}{L} - \frac{1}{CR}\right) \text{ respectively.}$$

Thus we have

$$\begin{aligned} e^{-\left(\frac{m \log_e \frac{n}{m}}{n-m}\right)} &\simeq e^{-\left(\frac{L}{CR^2} \log_e \frac{CR^2}{L}\right)} \\ &\simeq e^{\left(\frac{L}{CR^2} \log_e \frac{L}{CR^2}\right)} \simeq e^{(x \log_e x)} \simeq x^x, \end{aligned}$$

where $x = \left(\frac{L}{CR^2}\right)$, and using the identity $e^{\log_e y} = y$.

Similarly, we have

$$e^{-\left(\frac{n \log_e \frac{n}{m}}{n-m}\right)} \simeq x^{(1-x)}.$$

Further, $\left(\frac{n+m}{n-m}\right) = \left(\frac{1}{1-2x}\right)$ and thus $\eta = \left(\frac{x^x - x^{(1-x)}}{1-2x}\right)$

as a first approximation. Expanding in a power series we have

$$\eta = \left(\frac{1+x \log_e x - x}{1-2x}\right) = \left(\frac{(1-x)+x \log_e x}{1-2x}\right); \text{ where } x = \left(\frac{L}{CR^2}\right).$$

It will be clear that this expression for η is very much easier to compute, for particular values of L , C and R than the initial expression for η .

Applications of these Approximate Formulæ for t_1 and η , to a Typical Case.

In order to obtain an example of the type of accuracy to be expected in the use of these formulæ, and of the approximate magnitudes involved, the formulæ were applied to an impulse generator, using the normal Marx circuit, installed at Liverpool University. The wave front times (t_1) and stage efficiencies (η) were calculated for a number of values of generator resistance per stage (R), using both the exact formulæ and the simplified forms which the author has obtained. These values are tabulated in the table.

The critical stage resistance (given by $R = \sqrt{\frac{4L}{C}}$) for this generator, below which, in principle, the discharge ceases to be exponential and becomes oscillatory, is $R = 34.9$ ohms. It will be seen that the agreement between the values of η given by the exact and the approximate formulæ is exceedingly close for values of resistance as near the critical as 55 ohms, and for all higher values. The discrepancy between the two values of t_1 is at first rather wider, but above a stage resistance of

100 ohms this discrepancy too falls, and the agreement at all higher resistances is very close.

For high values of stage resistance it becomes impossible to use the more exact formulæ (without using 7-figure tables, etc.) because the quantity "*m*," a leading factor in these formulæ, is tending to zero, and cannot be exactly computed. The approximate formulæ are, however, accurate to almost any limit, under the same conditions as those which make it impossible to use the exact formulæ.

The wave tail time, that is, the time to *half* voltage on the wave tail, is calculated in each case by the approximate formula $t_2 = CR \log_e 2$ in order to show the type of impulse wave to which each value of resistance corresponds. It will be observed that the wave front time,

TABLE.

L (fixed). micro- henrys.	C (fixed). micro- farads.	R (vari- able) ohms.	t_1 (approx.). micro- secs.	t_1 (exact). micro- secs.	$\eta\%$ (approx.).	$\eta\%$ (exact).	t_2 (approx.). micro- secs.
61.0	0.2	40	2.53	3.34	81.8	76.8	5.55
61.0	0.2	55	2.54	2.95	83.5	83.5	7.63
61.0	0.2	70	2.42	2.66	87.3	87.5	9.71
61.0	0.2	100	2.13	2.23	91.9	92.1	13.87
61.0	0.2	150	1.75	1.77	95.3	95.3	20.8
61.0	0.2	250	1.30	1.31	97.9	97.9	34.7
61.0	0.2	330	0.785	—	99.25	—	73.5

whilst dependent on resistance, only varies over a much smaller proportionate range than the variation of resistance, whereas the wave tail time is proportional to the resistance.

Note on the Wave Tail Time (see fig. 3).

The value of the wave tail time is only exactly equal to $CR \log_e 2$ where the inductance *L* is zero, and the impulse wave then consists only of one exponential $v = Ve^{-mt}$, *m* at this limit being exactly $(1/CR)$. This ideal value $CR \log_e 2$ is shown as FA.

If we define t_2 as the time to half voltage on the wave tail and accordingly take a voltage $\left(\frac{\eta V}{2}\right)$ at the point C on the actual impulse wave, it becomes apparent that the condition of equality between $EC(=t_2)$ and FA ($=CR \log_e 2$) is that DE=BC.

If, however, these last two quantities are in some degree unequal EB is considerably greater than either, and $t_2(=EB+BC)$ will still approximate to $CR \log_e 2 (=EB+DE)$.

If we join AC, it is clear that this is very closely the tangent to the curve $v = Ve^{-mt}$, and therefore $BC = AB \cot \theta$, where

$$\tan \theta = -\frac{d}{dt}(Ve^{-mt}), \text{ for the value } t = CR \log_e 2 \text{ and } m = (1/CR),$$

that is,

$$\tan \theta = (V/2CR)$$

$$\text{and } BC = AB \cot \theta = FD \cot \theta = \left(\frac{V}{2}\right)(1-\eta) \times \frac{2CR}{V} = CR(1-\eta).$$

If BC is calculated from this expression for the typical impulse generator detailed above, values of BC about 20 per cent. greater than $t_1 (=DE)$ are obtained. In other words BC and DE , whilst not equal, are closely of the same order. But $t_1 (=DE)$ will rarely exceed 20 per cent. of t_2 and thus a difference of 20 per cent. between DE and BC will mean a difference of only 4 per cent. between $CR \log_e 2 (=DE + EB)$ and $t_2 (=BC + EB)$. Hence for most practical purposes the expression $CR \log_e 2$ has been shown to be a sufficiently exact value for the wave tail time, and the error will, in fact, usually be much less than 4 per cent. : that is, it is comparable with the errors in practical measurements.

Note on the Practical Application of the Above Theory.

Both the wave tail and the wave front times are modified by the load "resistance," or the resistance potential divider used for measurement. The stage efficiency will be higher than the generator efficiency as the stages may not easily be exactly synchronized by simultaneous spark gap breakdowns.

Further, the exact resistance per stage is never known, since the gap must present some resistance to the circuit, and there is often an additional resistance deliberately inserted. But the problem as stated above, although an idealized problem, is also an actual problem; the stage efficiency deduced being, in fact, the highest attainable efficiency, which theoretically might be obtained but could never be exceeded by any actual generator. These values obtained from the formulæ above represent convenient touchstones with which the behaviour of actual generators can be compared.

It may be added that the actual resistances per stage used in this generator vary between 550 and 35 ohms.

The type of mathematical analysis above could, it might be added, be applied to any physical problem whose mathematical form consisted of the difference between two exponential functions of widely different negative indices.

This work was mainly done while the author was a member of Professor Marchant's staff at Liverpool University, and he is indebted to Professor Marchant for the data of the actual generator installed there,

XL. *Notices respecting New Books.*

Photographic Optics (*The Manual of Photo-technique Series*). By ARTHUR COX, B.A., B.Sc. [Pp. 336.] (The Focal Press: London and New York, 1943. Price 15s.)

THIS book is one of a useful series of monographs, each dealing with one branch of the general subject. The volumes on development and enlarging have already appeared. This volume has the war-time merit of a certain austerity. The author writes for photographers with the object of explaining the action of the camera lens, and he has refused to be deflected by any academic irrelevancies. For example, the results of the mathematical treatment of the aberrations are utilized, but the mathematics is omitted. Again, the discussion of the resolution produced by a lens at points near the axis is limited to what is practically relevant in view of the power of a normal photographic emulsion to render detail.

This emphasis on the practical point of view has resulted in a handbook of outstanding quality on a section of the subject which has been badly neglected in the past. The subject-matter is not original but the line of approach is. The author has collected in one volume a mass of data previously out of reach of the majority of photographers, and the book will undoubtedly be found a valuable addition to the photographer's library. B. D. H. W.

ADDENDUM.

Corrections to paper on "The Force required to give a Small Acceleration to a Slowly-moving Sphere carrying a Surface Charge of Electricity." By G. F. C. SEARLE, *Philosophical Magazine*, ser. 7, vol. xxxiii. December 1942.

Page 896, line 2 of §4, for " $t =$ " read " $t=0$."

Page 898, line 10,* for dW_u/dt read dW_u/du .

line 11,* for $\frac{f}{u} \frac{dW_u}{dt}$ read $\frac{f}{u} \frac{dW_u}{du}$.

The last formula on page 898—correct in the "slip proof"—has suffered some type accidents. It should read

$$m_i = \frac{\mu q^2}{a} \left\{ \frac{2}{3} + \frac{4u^2}{5v^2} + \frac{6u^4}{7v^4} + \dots \right\}.$$

* Author's errors.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XLI. *Standard Curves.*

By C. W. HANSEL, Bedford School *.

[Received August 6, 1942.]

THE cartesian plot of $y=f(x)$ is usually obtained by calculating y for convenient values of x and then plotting the corresponding values of y against x using rectangular axes. Each function of x gives its own particular curve. The real roots of $f(x)=0$ are found at the intersections of the curve with the axis of x . More accurate values of the roots may then be found by well-known methods.

An alternative method is to use one standard curve to represent all functions of the same type, and to adjust the scales until this curve represents the cartesian plot of any selected function of that type. For example, the curve $Y=X^2$ may be used as a standard curve to represent any quadratic function such as $y=ax^2+bx+c$. By using $Y=X^2-2X$ as the standard curve, a change of origin is avoided. Similarly, using the standard curves $Y=X^3-3X^2$, $Y=\sin X$, and $Y=e^X$, $Y=e^{-X^2}$, any cubic, sine, or exponential function may be represented by a suitable change of scales without any transformation of origin.

1. *Notation.*

$Y=F(X)$ is the equation of a standard curve plotted to rectangular co-ordinates OX , OY the origin being O .

$[X]$ denotes the figured or graduated axis of X and is read "scale X ."

Similarly, $[Y]$ is read "scale Y ."

A graduation or scale number a on $[X]$ (whether figured or not) is denoted by $a[X]$, which is read " a on X ."

Actual points are denoted by capitals. " $B(x, y)$ " means "a point B having co-ordinates x, y ."

" $A : B$ " means "the join of A and B or AB ."

" $A : B \rightarrow C$ " means that "the join AB passes through C ."

" $(A : B ; C : D) \rightarrow E$ " means that " AB intersects CD at E ," which may also be written $(AB ; CD) \rightarrow E$."

A *secor* is a straight line connecting two points or passing through a pivot point which intersects a curve or a straight line at a required point or points. The cursor line of a slide rule or the transversal of a nomogram is a *secor*. It is the "ablesegerade" or reading-off point of the German literature. It may be a rotatable line about a pivot point or one of the rays of a projection pencil. The function of a *secor* is to establish a numerical relation between the points collinear with it.

* Communicated by the Author.

" $a[x] : b[y] \rightarrow c[z]$ " means "the join of a on scale x and b on scale y passes through c on scale z " or "a secor connecting $a[x]$ and $b[y]$ passes through $c[z]$."

Let $[X]$ be a primary scale on an axis OX , and $[x]$ a secondary scale along ox with corresponding scale numbers k times those of $[X]$. This is described as "applying a scale factor k to $[X]$ or $[x] = [kX]$."

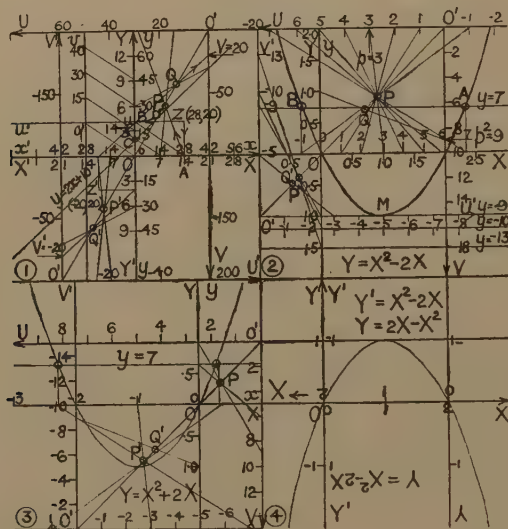
$a[1] \equiv a^2[2]$ means "the graduation mark figured a on scale 1 is opposite the graduation mark figured a^2 on scale 2," and reads " a on 1 is opposite a^2 on 2."

$a[1] \nearrow a[5]$ means "a reading a on scale 1 is transferred to the same reading on scale 5" and reads "transfer a from 1 to 5."

2. Transformation of Scales.

In 1.1, $[X]$ and $[x]$ are adjacent scales having the same graduation marks, but the scale numbers on $[x]$ are double those on $[X]$. If $[X]$ is

Fig. 1.



the primary scale, the scale factor k for transforming $[X]$ into $[x]$ is 2. $[x] = [2X]$. Similarly, $[y] = [5Y]$.

$O'U$, $O'V$ are auxiliary axes passing through any desired origin and figured in any desired manner.

$[U]$ is figured in the opposite direction to $[x]$ or $[X]$

$[V]$ is figured in the opposite direction to $[y]$ or $[Y]$.

It is required to transform $[x]$ into the more convenient scale $[U]$.

Choose equal numbers a on the two scales and connect them by a straight line. $a[x] : a[U] \rightarrow P$ on OO' .

P is the centre of projection for transforming $[x]$ into $[U]$ or vice versa. Any secor through P cuts $[x]$ and $[U]$ at equal scale readings.

In 1, $14[x] : 14[U] \rightarrow P$ on OO' .

Similarly, $30[y] : 30[V] \rightarrow Q$ on OO' . Q is the centre of projection for the transformation of $[y]$ into $[V]$.

To find the x , y and U , V co-ordinates of any point Z .

Draw $ZA \perp OX$, then AP meets $[U]$ in $U=x$.

Draw $ZB \perp OY$, then BQ meets $[V]$ in $V=y$.

This operation may be denoted by $Z(x, y=U, V)$ which is read "finding the x , y or U , V co-ordinates of Z ."

For negative co-ordinates the auxiliary scales $[U']$ and $[V']$ are used and the centres of projection P' and Q' .

$Z(x, y=U, V)$ gives $x=U=28, y=V=20$.

$Z'(x, y=U, V)$ gives $x=U'=-20, y=V'=-20$.

To plot $y=2x+10$ to $[x]$ and $[y]$, plot the points $C(0, 10)$ and $D(14, 38)$. $C:D$ is $y=2x+10$.

If the centre of projection is at infinity, the rays of the projection pencil become parallel.

1.1 shows the parallel projection of $[x] \rightarrow [u]$ and of $[y]$ into $[v]$. In this case, there is a transformation of origin from o to $o'(-30, 20)$ and $[u]=[x]$ and $[v]=[y]$.

3. Transformation of Quadratic Functions.

$$Y = X^2 - 2X \rightarrow y = x^2 - 2px.$$

The curve $Y = X^2 - 2X$ passes through O, O and $2, O$. The minimum point M of the curve is $1, -1$ and the minimum value of $Y(m) = -1$.

The curve $y = x^2 - 2px$ passes through O, O and $2p, O$. M is the point $p, -p^2$ and $m = -p^2$.

Hence, $Y = X^2 - 2X$ and $y = x^2 - 2px$ coincide if $[x] \rightarrow [pX]$ and

$$[y] \rightarrow [+p^2Y].$$

1.2 illustrates the transformation

$$Y = X^2 - 2X \rightarrow y = x^2 - 6x. \quad p=3; \quad m = -p^2 = -9.$$

Choose any desired auxiliary scales $[U], [V], [U'], [V']$. Find the centres of projection, P, Q , for positive values, and P', Q' for negative values

$$3[U] : 1[X] \rightarrow P \text{ on } OO'; \quad -3[U'] : -1[X] \rightarrow P' \text{ on } OO'.$$

$$9[V] : 1[Y] \rightarrow Q \text{ on } OO'; \quad -9[V'] : -1[Y] \rightarrow Q' \text{ on } OO'.$$

$[U]$ may now be projected along ox and $[V]$ along oy , but in general, it is easier and more accurate to use $[U]$ and $[V]$ rather than $[x]$ and $[y]$.

The line $y=7$ cuts $y=x^2-6x$ at A where $x=7$, and B where $x=-1$. These values of x are the values of the roots of $x^2-6x-7=0$.

The equation $x^2-6.28x-7.35=0$ is no more difficult to solve than the previous equation, but it is more convenient to work with the auxiliary scales instead of $[x]$ and $[y]$.

If a plot of $y=x^2-6x-7$ is required, the origin must be moved and $[y]$

must be renumbered or projected from [U] using a new centre of projection on $o'o'$.

$$o \rightarrow o'(0, 7).$$

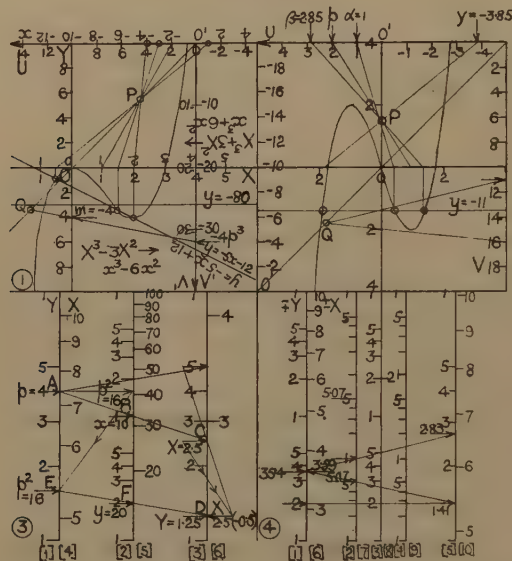
If $x^2 - 2px + q = 0$ has $q > p^2$, the roots are imaginary and equal to $p \pm j\sqrt{q - p^2}$. $q - p^2$ is the distance from M to the line $y = -q$.

Hence, $x^2 - 6x + 10 = 0$ has roots $3 \pm j$,
and $x^2 - 6x + 13 = 0$ has roots $3 \pm 2j$.

Scales for $\sqrt{q - p^2}$ may be marked adjacent to [y] and [V] so that the imaginary roots may be read off easily.

1.3 shows that $Y = X^2 - 2X \rightarrow Y = X^2 + 2X$ by a transformation of scales. The construction shows that the roots of $x^2 + 6x - 7 = 0$ are 7 and -1 .

Fig. 2.



It is also apparent from 1.3 that 1.2 can be used instead of 1.3 if the numbers of x change sign, since $Y = X^2 - 2X \rightarrow Y = X^2 + 2X$ if $X \rightarrow -X$. Hence, to change 1.2 to 1.3 a scale factor of -1 must be applied to all the horizontal scales.

1.4 shows that $Y = 2X - X^2 \rightarrow Y' = X^2 - 2X$ either by reversing [Y], i. e., $[Y] \rightarrow [-Y]$ or by inverting the curve and renumbering the scales.

4. Transformation of Cubic Functions.

$$Y = X^3 - 3X^2 \rightarrow y = x^3 - 3px^2.$$

If M and M' are the minimum and maximum points on the curve,

$$M \text{ is } 2p, -4p^3; m = -4p^3; M' \text{ is } 0, 0; m' = 0.$$

2.1 shows the transformation $Y = X^3 - 3X^2 \rightarrow y = x^3 - 6x^2$.

In this case $p = 2$; $m = -32$. Auxiliary scales [U] and [V] are used.

The line $y = -5x - 12$ cuts the curve where $U = 3, 4$, and -1 . These are the roots of the equation $x^3 - 6x^2 + 5x + 12 = 0$.

Since both scales $[y]$ and $[V']$ lie beyond the centre of projection Q , these scales are numbered in the same direction.

For $x^3 - 9x^2 + 80 = 0$, $p = 3$; $m = -108$.

To use 2.1, apply a scale factor $k = \frac{3}{2}$ to $[U]$.

The line $y = -80$ or $Y = -4 \times \frac{80}{108}$ cuts the curve where $U = 4, 7.63$, and -2.63 . These are the roots of $x^3 - 9x^2 + 80 = 0$.

(The solution is not shown in 2.1; only the line $y = -80$.)

The transformation $Y = X^3 - 3X^2 \rightarrow Y' = X^3 + 3X^2$ is made by reversing the sign of all numbers on $[X]$ and $[Y]$ since $Y \rightarrow -Y'$, if $X \rightarrow -X$.

The curve $y = x^3 + 3px^2$ passes through $0, 0$ and $-3p, 0$. M is $0, 0$ and $m = 0$; M' is $-2p, 4p^3$ and $m' = 4p^3$.

If 2.1 is inverted, it shows the transformation

$$Y = X^3 - 3X^2 \rightarrow Y = X^3 + 3X^2.$$

Inverting 2.1, it will be seen that the straight line $y = -5x + 12$ cuts $y = x^3 + 6x^2$ at $x = 1, -3$, and -4 , which are the roots of

$$x^3 + 6x^2 + 5x - 12 = 0.$$

2.2 shows the transformation $Y = X^3 - 3X \rightarrow y = x^3 - 3p^2x$.

M is $p, -2p^3$; $m = -2p^3$; M' is $-p, 2p^3$; $m' = 2p^3$.

The curve $y = x^3 - 12x$ is cut by the straight line $y = -11$ where $x = 1, 2.85$, and -3.85 .

The transformation $Y = X^3 \rightarrow y = ax^3$ may be made using scale factors only.

$y = x^3$ cuts $y = 6x - 9$ where $x = -3$. The remaining roots are given by the quadratic equation $x^2 - 3x + 3 = 0$ and are $1.5 \pm 0.866j$.

$y = 0.5714x^3$ cuts $y = 0.7143x - 0.2857$ where $x = 0.50, 0.78$, and -1.28 , the roots of the cubic equation $0.5714x^3 - 0.7143x + 0.2857 = 0$.

The previous transformations enable the real roots of any cubic equation to be found without using the Cardan transformation whereby the term in x^2 is eliminated. If the equation contains no term in x^2 it is unnecessary to make the coefficient of x^3 equal to unity.

To find the imaginary roots of $x^3 + qx^2 + rx + s = 0$. First find the real root α .

Let $a \pm j.b$ be the imaginary roots.

Then $2a + \alpha = -q$; $a = -(\alpha + q)/2$.

b may be obtained from $a^2 + b^2 = -s/\alpha$ after a has been found. It is easy to construct a nomogram or a slide rule for the rapid evaluation of b .

5. Derived Curves.

If $Y = X^3 - 3X^2$; then $D(Y)$ or $Y' = 3X^2 - 6X = 3(X^2 - 2X)$.

If $y = x^3 - 3px^2$; then $D(y)$ or $y' = 3x^2 - 6px = 3(x^2 - 2px)$.

3.3 shows the standard curves $Y=X^3-3X^2$ and $Z=X^2-2X$.

These curves may be transformed into $y=x^3-3px^2$ and $y'=3(x^2-2px)$ by a transformation of scales.

Example. (See 3.3.)

$$Y=X^3-3X^2 \text{ (Curve I.)} \rightarrow y=x^3-6x^2. \quad p=2; \quad m=-32.$$

The centres of projection P and Q' are found

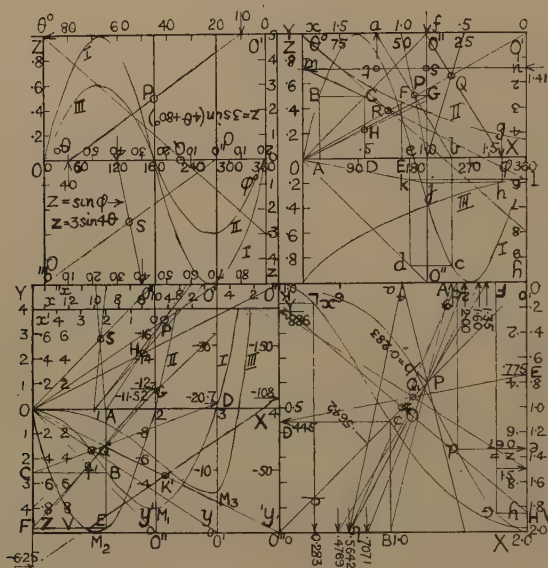
$$2[x]:1[X] \rightarrow P \text{ on } 0'0; \quad -32[y]:-4[Y] \rightarrow Q' \text{ on } 00'.$$

To find y and y' if $x=2.4$.

$$2.4[x] \rightarrow P \rightarrow A[X] \rightarrow B(I) \rightarrow C(Y) \rightarrow Q' \rightarrow D \text{ or } -20.7[y].$$

When $x=2.4, \quad x^3-6x^2=-20.7.$

Fig. 3.



Again, $Z=X^2-2X \rightarrow y'=3x^2-12x=3(x^2-4x).$

To find y' when $x=2.4$.

$$2[x']:1[X] \rightarrow S[0'0]; \quad -12[y']:-1[Z] \rightarrow T'[0'0].$$

$$2.4[x'] \rightarrow S \rightarrow A[X] \rightarrow E(II) \rightarrow F[Z] \rightarrow T' \rightarrow \\ \rightarrow G[y']. \quad G[y']=-11.52.$$

Thus, if $y=x^3-3px^2$ it is possible to find y and dy/dx for any value of x .

Again, if $y=x^3-3px^2+qx+r$ it is possible to obtain the curve for y' graphically with considerable accuracy. The curve for y may be obtained by adding $qx+r$ to the curve obtained by projection.

For values of x for which y is positive, the standard curves must be drawn for the desired range of values of x and y required.

If $Y=X^3-3X^2$; then $D^{-1}Y$ or $'Y=X^4/4-X^3$.

If $y = x^3 - 3px^2$; then $D^{-1}y$ or $'y = x^4/4 - px^3$.

$'y$ may be read "dash y " and means $\int y dx$.

In 3.3, Curve III. is $'Y$ or $V = X^4/4 - X^3$.

The minimum point M_3 is 3, $-27/4$.

For $'Y = X^4/4 - pX^3$ the minimum point is $3p$, $-27p^4/4$.

For $y = x^3 - 6x^2$; $'y = x^4/4 - 2x^3$; $p = 2$.

Hence, M_3 is 6, -108 .

The centres of projection H and K' are found

$$6[x] : 1[X] \rightarrow H \text{ on } OO'''; -108[y] : -6.25[Y] \rightarrow K' \text{ on } OO''.$$

If $y = x^3 - 3px^2$, the value of $'y$ for any value of x may be found, and the integral curve may be obtained graphically by projection of the standard curve. Any arbitrary constant is easily dealt with by renumbering the $'y$ scale, or by taking a new $'y$ scale at a convenient point along OO''' as origin.

The following curves are not shown:—

If $Y = X^3 - 3pX^2$; M_1 is $-2p$, $4p^3$. (Curve IV.)

$Y' = 3(X^2 + 2pX)$; M_2 is $-p$, $-p^2$. (Curve V.)

$'Y = X^4/4 + pX^3$; M_1 is $-3p$, $-27p^4/4$. (Curve VI.)

Curve IV. is the curve $Y = X^3 - 3pX^2$ inverted, or the mirror image of Curve I. in the axis of Y (see 2.3 and section 4).

Curve V. is the mirror image of $Y = 3(X^2 - 2pX)$ in the axis of z (see section 3).

Curve VI. is the mirror image of $Y = X^4/4 - pX^3$ in the axis of V .

6. Nomographic Transformations.

The graphical processes already described suggest similar nomographic processes. It will be shown in a future paper that logarithmic scales may be so disposed in a nomogram that four to five-figure accuracy may be obtained on a drawing measuring about $30'' \times 25''$. For this reason logarithmic scales have been preferred to other possible scales in the construction of nomograms.

Consider the transformation

$$Y = X^2 - 2X \rightarrow y = x^2 - 2px.$$

$$x = pX; \log x = \log p + \log X.$$

$$y = p^2Y; \log y = \log p^2 + \log Y.$$

Nomogram 2.3 consists of logarithmic scales [1], [2], [3], and three functional scales [4], [5], [6] each figured X and graduated so that the adjacent scale gives $Y \equiv X$. (It may be necessary to apply a scale factor 10^n to $[Y]$.)

Thus, $X[4] \equiv 10Y[1]; X[5] \equiv 10^2Y[2]; X[6] \equiv Y[3].$

The transformations are illustrated best by means of examples (see 2.3).

Example 1. Find $y=x^2-8x$ when $x=10$.

$$p=4. \quad (p \text{ or } 4 [1]) \equiv (p^2 \text{ or } 16 [2]).$$

$$(p=4 [1] \text{ or } A) : (x=10 [2] \text{ or } B) \rightarrow (X=2.5 \text{ or } -0.5 [3] \text{ or } C).$$

By inspection $X=2.5$.

$$2.5 [6] \equiv (Y=1.25 [3] \text{ or } D) : (p^2=16 [1] \text{ or } E) \rightarrow (y=20 [2] \text{ or } F).$$

Example 2. If $x^2-8x=20$, find x .

$$p=4. \quad (p^2=16 [2]) \equiv p [1].$$

$$(20 [2] \text{ or } F) : (E \text{ or } p^2=16 [1]) \rightarrow (Y=1.25 [3] \text{ or } D) \equiv (X=2.5 \text{ or } -0.5 [6]).$$

$$(X=2.5 [6]) \nearrow (X=2.5 [3] \text{ or } C) : (A \text{ or } p=4 [1]) \rightarrow (x=10 [2] \text{ or } B).$$

Similarly, $X=-0.5$ (disregard the $-$ sign) gives $x=2 [2]$.

Example 3. If $x^2-1.2x+5.2=0$, find x .

$$p=0.6 \quad (p=0.6 [1]) \equiv 0.36 [2].$$

$p^2=0.36$ and is less than $q=5.2$. Hence the roots of the equation are imaginary and equal to $p \pm j\sqrt{q-p^2}$ or $0.6 \pm j \cdot \sqrt{4.84}$. $4.84 [2] \equiv 2.2 [1]$.

Hence, the roots of the equation are $0.6 \pm j \cdot 2.2$.

It is simpler to add or subtract mentally than to use a nomogram.

If the given quadratic is of the form $ax^2+bx+c=0$, it must be reduced to the form $x^2-2px+q=0$. $p=2b/a$ and $q=c/a$ are easily obtained using the logarithmic scales.

2.4 illustrates the nomographic transformation

$$Y=X^3-3X \rightleftarrows y=x^3-3p^2x.$$

[1], [2], [3], [4], [5] are logarithmic scales.

[6], [7], [8], [9], [10] are graduated so that $\pm X$ on any one of these scales is opposite $\mp Y$ on the adjacent scale.

$$p^2 [1] \equiv p^3 [2] \equiv p^2 [3] \equiv p^3 [4] \equiv p [5].$$

The transformations are best illustrated by means of examples.

Example 1. Find $y=x^3-6x$, if $x=5.07$.

$$p^2=2. \quad 2 [1] \equiv 2 [3] \equiv 1.41 [5] \equiv 2.83 [2] \equiv 2.83 [4].$$

$$x=pX \text{ gives } x. \quad 5.07 [2] : 1.41 [5] \rightarrow X=3.59 [6] \equiv (Y=35.4 [1]).$$

$$y=p^3Y \text{ gives } y. \quad 35.4 [1] : 2.83 [5] \rightarrow (y=100 [2]).$$

Example 2. If $y=x^3-6x=100$, find x .

$$p=2. \quad 2 [1] \equiv (p^3=2.83 [2]) \nearrow (2.83 [5]).$$

$$y=p^3Y \text{ gives } Y. \quad 2.83 [5] : 100 [2] \rightarrow (Y=35.4 [1]) \equiv (X=3.59 [6]).$$

$$x=pX \text{ gives } x. \quad 3.59 [6] : 1.41 [5] \rightarrow x=5.07 [2].$$

Thus the nomogram 2.4 gives one real root α of any cubic equation of the form $x^3-3p^2x=r$. The remaining roots may be found by solving the quadratic obtained by dividing $x^3-3p^2x-r=0$ by $x-\alpha$.

More accurate values of the roots may be found by successive approximation.

6a. Harmonic Curves.

3.1 illustrates $Z = \sin \phi \rightarrow z = 3 \sin 4\theta \rightarrow$

$$\rightarrow z = 3 \sin (4\theta + 80^\circ) \text{ or } z = 3 \sin 4(\theta + 20^\circ).$$

Curve I. is the standard curve $Z = \sin \phi$.

Curve II. is the curve $z = 3 \sin 4\theta$ to axes through O' .

Choose convenient auxiliary scales for θ and z .

$$10^\circ [\theta] : 40^\circ [\phi] \rightarrow P ; 3 [z] : 1 [Z] \rightarrow Q.$$

Curve II. is obtained by projection of $Z = \sin \phi$ using P and Q .

The transformation $z = 3 \sin 4\theta \rightarrow z = 3 \sin 4(\theta + 20^\circ)$, involves a change of origin from $\theta = 0^\circ$ to $\theta = 20^\circ$. The new z and θ scales are renumbered from O'' . If O'' is situated inconveniently on the θ scale, mark a convenient θ scale numbered from O''' and project through S to this scale. S lies on $O''O'''$ and is easily found using two equal values of θ .

The transformation $Z = \cos \phi \rightarrow z = a \cos p\theta \rightarrow z = a \cos (p\theta + \alpha)$ is similar to the sine transformation.

Scales for θ should be marked in degrees and radians.

7. Damped Harmonic Curve.

The standard curves $Y = e^{-x}$ and $Z = \sin \phi$ may be used to carry out the transformation

$$(Y = e^{-x})(Z = \sin \phi) \rightarrow y = e^{-x} \cdot \sin \theta.$$

This involves the graphic multiplication of two ordinates. For example, to find $y = e^{-0.8} \cdot \sin 30^\circ$.

In 3.2, OO'' is drawn at 45° with the axes.

A is $30^\circ [\phi]$. $AB = \sin \phi$. Draw BC to meet OO'' .

Draw CD at right angles to the axis of ϕ . Then $OD = \sin \cdot 30^\circ$.

E is $0.8 [X]$. $EF = e^{-0.8}$. Draw FG parallel to OX .

Join OG cutting CD in H . Then $DH = OG \cdot OD = e^{-0.8} \cdot \sin 30^\circ$.

3.2 also shows the transformation

$$(Y = e^{-x})(Z = \sin \phi) \rightarrow (v = Ae^{-qx})(Z = B \cdot \sin \cdot p\theta) \\ \rightarrow y = R \cdot e^{-qx} \sin \cdot p\theta \text{ where } R = A \times B.$$

Example.—To find $y = v \cdot z = 4e^{-2x} \cdot 2 \cdot \sin \cdot 4\theta$, when $\theta = 60^\circ$, and $x = 0.8$.

a is $60^\circ [\theta] \rightarrow P \rightarrow b \rightarrow c$ (Curve I.) $\rightarrow d(OO'') \rightarrow e [\phi]$.

$Oe = \sin \cdot 4\theta = \sin \cdot 240^\circ$.

f is $0.8 [x] \rightarrow Q \rightarrow g [X] \rightarrow h$ (Curve III.) $\rightarrow j (O'O'')$.

Oj cuts ed at $k \rightarrow l [y] \rightarrow R \rightarrow m [Y] \rightarrow n [y]$.

$O'n$ is $y = 1.41$.

If the angle θ is in radians, a scale of ϕ radians may be marked along the axis of ϕ as well as a scale of degrees.

If 3.2 be inverted, the point s is x , $y = 0.8, 1.41$, and the point t is θ° , $y = 60^\circ, 1.41$.

The complete curve for a damped simple harmonic motion may be obtained by plotting a series of points in this way.

8. Normal Variation and Random Error.

The frequency, y , of random errors or deviations of magnitude x is given by $y = ke^{-p^2x^2}$.

The probability, Z , of an error of magnitude $+x$ and $-x$ is given by $Z = 2p/\sqrt{\pi}(e^{-p^2x^2})$.

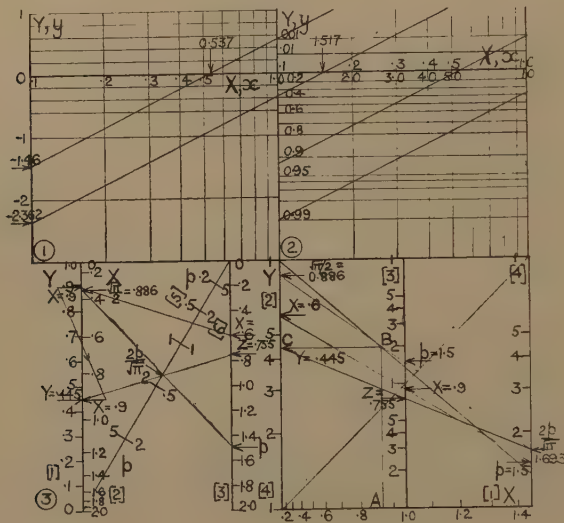
3.4 shows the standard error curve $Y = e^{-X^2}$ and the transformations

$$Y = e^{-X^2} \rightarrow y = e^{-p^2x^2}; \quad Y = e^{-X^2} \rightarrow Z = 2p/\sqrt{\pi}(e^{-p^2x^2}).$$

The transformations are illustrated by the following example (3.4):—
Find $y = e^{-(1.5)^2(0.6)^2}$ and Z .

$$p = 1.5; \quad x = 0.6.$$

Fig. 4.



Choose any arbitrary scales $O'X'$, $O'V$.

$1.0[X'] : 1.5[X] \rightarrow P$ (the centre of projection for $[X]$ and $[X']$).

(A or $x = 0.6[X']$) $\rightarrow P \rightarrow B[X] \rightarrow C$ (on curve) $\rightarrow D$.

$$D \text{ is } y = e^{-(1.5)^2(0.6)^2} = 0.445.$$

Find Q on OO' so that $V = 2p/\sqrt{\pi}(Y)$. $\sqrt{\pi}/2 = 0.886$.

$1.5[V] : 0.886[Y] \rightarrow Q_1$ on OO' . $D \rightarrow Q_1 \rightarrow Z[V]$; $Z = 0.775$.

A scale may be marked along OO' for $2p/\sqrt{\pi}$ and figured p , each value of p giving the corresponding centre of projection Q_1 (see 3.4).

4.3 is a nomogram for working the previous example. [1] is a uniform scale for Y . [2] gives the corresponding value of X . [4] is a uniform scale negatively figured. [5] is a segmentary scale. [6] is the reciprocal scale of [5].

$$0.6 [4] : (p=1.5 [6]) \rightarrow px=0.9 [1] ; 0.9 [1] \nearrow 0.9 [2],$$

$$0.9 [2] \equiv (Y=0.445 [1]). \quad y=e^{-p^2x^2}=e^{-(1.5)^2(0.6)^2}=0.445.$$

$$\sqrt{\pi}/2 [1] : p [4] \rightarrow 2p/\sqrt{\pi} [5] ; 2p/\sqrt{\pi}=1.69.$$

$$0.445 [1] : 1.69 [5] \rightarrow (Z=0.755 [4]).$$

If [3] is figured with values of Y corresponding to X [4], then

$$0.6 [1] : 1.5 [5] \rightarrow X [4] \text{ and } Y [3].$$

4.4 shows the straight line obtained by plotting $\log . Y$ (figured Y) against X (figured X). There are two additional logarithmic scales, [3] and [4].

$$0.6 [2] : p [4] \rightarrow (X=0.9 [3]) \nearrow (0.9 [1] \text{ or } A).$$

$$A \rightarrow B \text{ (on line)} \rightarrow (C \text{ or } Y=0.445 [2]).$$

$$(0.886 \text{ or } \sqrt{\pi}/2 [2]) : (1.5 \text{ or } p [3]) \rightarrow (2p/\sqrt{\pi} \text{ or } 1.69 [4]).$$

$$1.69 [4] : 0.445 [2] \rightarrow Z=0.755 [3].$$

The following example indicates the application of the present method to statistical problems. The data are those quoted by G. C. Whipple for heights of soldiers in an army of 18,780 men (see 'The Theory of Measurements,' by Tuttle and Satterly):—

H inches.	C inches.	x =66.8 —C.	N soldiers.	u per cent.	$y = \frac{u}{16}$.	$\text{lolog}_{10} y$.	$\text{lolog}_{10} y$.	$\log_{10} x$.
60-61	60.5	6.3	197	1.05	0.0656	0.0730		0.7993
61-62	61.5	5.3	317	1.69	0.1056	1.9896	—0.0104	0.7243
62-63	62.5	4.3	692	3.69	0.2306	1.8043	—0.1957	0.6335
63-64	63.5	3.3	1,289	6.86	0.4287	1.5657	—0.4343	0.5185
64-65	64.5	2.3	1,961	10.44	0.6527	1.2678	—0.7322	0.3617
65-66	65.5	1.3	2,613	13.91	0.8694	2.7838	—1.2162	0.1139
66-67	66.5	.3	2,974	15.84	0.9900	3.6356	—2.3644	1.4771
67-68	67.5	—0.7	3,017	16.07	1.0040	3.2390	—2.7610	1.8451
68-69	68.5	—1.7	2,287	12.18	0.7612	1.0737	—0.9263	0.2304
69-70	69.5	—2.7	1,599	8.52	0.5325	1.4372	—0.5628	0.4314
70-71	70.5	—3.7	878	4.67	0.2918	1.7283	—0.2717	0.5682
71-72	71.5	—4.7	520	2.77	0.1732	1.8817	—0.1184	0.6721
72-73	72.5	—5.7	262	1.39	0.0869	0.0258	—	0.7559
73-74	73.5	—6.7	174	0.92	0.0575	0.0935	—	0.8261
			18,780	100.00	6.25			

H is the range of heights of soldiers in each group.

The mean height \bar{H} is 66.8".

C is the central height of each group (or cell).

N is the number of soldiers in each group.

x is the deviation from the mean height in each group.

u is the percentage of soldiers in each group. If u is plotted against x , the maximum ordinate of the curve is 16.0 for $x=0$.

$y=u/16$. y plotted against x gives a curve closely approximating to the curve of normal error, deviations being most marked for small values of y .

3.4 shows the positive half of the standard curve $Y=e^{-X^2}$ and the curve $y=e^{-p^2x^2}$, drawn inverted to an origin at O' . With 3.4 inverted, $1.00[Y] : 1.00[y] \rightarrow Q_2$ on OO' . (This sector is not drawn.)

If G is a point on curve y for which $x=1$, $G \rightarrow H \rightarrow Q_2 \rightarrow K \rightarrow L$ gives the corresponding point L on the curve Y for which $X=p=0.283$. Other pairs of corresponding points may be found in a similar way and a value of p may be obtained for each pair. The value is 0.283 for $x=2, 3, 4$, but $x=5$ and 6 give values of p which deviate from this value.

Hence the curve approximates to $y=e^{-0.283^2x^2}$ and $u=16e^{-0.283^2x^2}$.

The dispersion, average deviation, and standard deviation are obtained by projecting $X=0.4769, 0.5642$, and 0.7071 into $x=1.35, 1.60$, and 2.00 , using the centre of projection P .

The probability of an error $+x$ and $-x$ (z) is obtained using the multiplying factor 0.5642 on OO' . For example, when $x=4.0$;

$$a \rightarrow b \rightarrow Q_2 \rightarrow c \rightarrow d \rightarrow (z=0.67).$$

$$\text{Also, } k=0.5642p \cdot dx, \quad p=0.16/0.5642=0.283.$$

4.1 shows the working of the problem, using a new type of probability paper. The scale for $\pm X$ and $\pm x$ is logarithmic, figuring positive and negative values of X and x from 0.1. The scale for Y and y is a lolog scale figuring lolog Y and lolog y from 0.00. The plotted values of x are one-tenth their actual values resulting in a 10-fold value of p .

The standard curve is

$$-\text{lolog} . Y = \text{lolog} . 0.4343 + 2 . \text{log} . X \text{ or } -\text{lolog} . Y = 2 . \text{log} . X - 0.3622.$$

This curve cuts $[Y]$ where $\text{lolog} . Y = -2.362$ and cuts $[X]$ where $X=1.517$.

$$\text{Also } -\text{lolog} . y = 2 . \text{log} . x - 0.3622 + 2 . \text{log} . p,$$

and cuts $[y]$ where

$$\text{lolog} . y = -1.46 = -2.3622 + 2 . \text{log} . p, \text{ and cuts } [x] \text{ where } x=0.537.$$

$$\text{Hence } 2 . \text{log} . p = 2.3622 - 1.46 = 0.9022, \text{ giving } p=2.83.$$

$$\text{Again, } p=1.517/0.537=2.83.$$

Hence the straight line represents $u=16 \cdot e^{-2.83^2x^2}$ and is parallel to the straight line for $Y=e^{-X^2}$.

4.2 shows the same type of solution using paper ruled logarithmically for X and x , and lologarithmically for Y and y but figured Y and y . Using this paper, the points may be plotted rapidly without referring to mathematical tables.

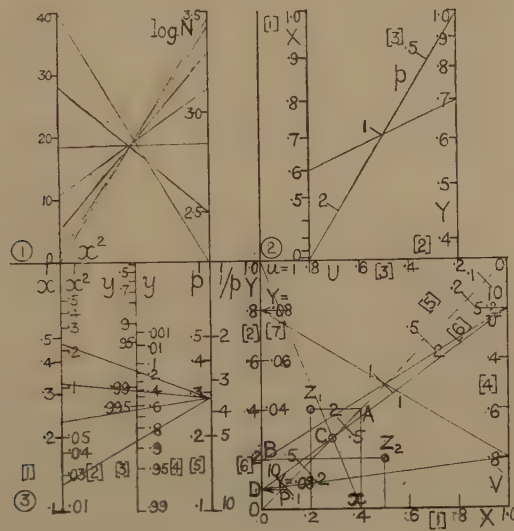
4.1 may be extended into 4.2 if the lower scale for $\pm X$ and $\pm x$ is used. For 4.2, the upper scale is appropriate.

9. Errors of Observation.

If errors of observation conform to the exponential law of errors, it is legitimate to apply the method of least squares in order to reduce the observations to their best values. It is therefore desirable to test the data before such a reduction is made. In a paper "An Extension of Nomography" (Phil. Mag. xxxiv. p. 1 (1943)), the author has indicated how the accuracy of observations may be tested on a nomogram, how to decide which observations to reject, and how a law may be found to fit the observations in a few cases without using the method of least squares. The methods there described may be used when the number and accuracy of the observations are insufficient.

The data already given in Section 8 may be used to illustrate how measurements or observed data may be studied with a view to discovering

Fig. 5.



how their deviations from an assumed law are consistent with the normal law of error.

5.1 illustrates a preliminary test which can be applied to the measurements without any elaborate or laborious preliminary reduction. Only the positive values of the independent variate x have been considered.

[1] is a uniform scale for x^2 . [2] is a reversed uniform scale for $\log_{10} N$. If $N = ke^{-p^2x^2}$, the joins $x^2[1] : \log N[2]$ are concurrent. The diagram shows that the exponential law of error is followed for values of x less than about 5, but there is a substantial deviation from the law for values of x having greater values. The law is not followed for values of x exceeding 5.

The method of finding k has already been indicated. It remains to find p as expeditiously as possible and to a degree of accuracy consistent

with the fit of the formula to the observations. p may be found using 4.2 (already described) or by using 5.2 or 5.3.

5.2 is a nomogram for the rapid determination of p . [1] is a square scale for X . [2] is a logarithmic scale to base e for Y . [3] is a segmentary scale for p^2 (figured p).

$$X [1] : Y [2] \rightarrow p [3].$$

Additional scales may be provided for other ranges of values of X and Y .

Owing to the closeness of the graduations near the zero of the square scale, it may be desirable to use 5.3 for small values of X .

5.3 is a nomogram for $Y = e^{-X^2} \rightarrow y = ke^{-p^2 x^2}$.

[1], [2], [5], [6] are logarithmic scales to base 10.

[3], [4] are lologarithmic scales to base 10.

Since $\text{lolog}_{10} (1/y) = \log 0.4343 + \log x^2 + \log p^2$, the scales may be figured so that

$$x [1] : p [5] \rightarrow y [3] ; 10x [1] : p [5] \rightarrow y [4].$$

$$x [1] : 10p [5] \rightarrow y [4] ; 10x [6] : p [1] \rightarrow y [3].$$

$$x [1] \equiv x^2 [2] ; p [5] \equiv 1/p [6].$$

If the observations conform to the exponential law of errors, and if the data are reduced so that the maximum value of $y=1$ when $x=0$ (as indicated in section 8), the joins $x [1] : y [3]$ should be concurrent at $p [5]$.

10. Product, Reciprocal, and Quotient Curves.

If Y_1, Y_2 are the ordinates of two standard curves, for the same or different values of X , the product curve is the curve whose ordinate is $Y = Y_1 Y_2$, the quotient curve is Y_1/Y_2 , and the reciprocal curve is $Y = 1/Y_1$.

5.4 illustrates a well-known method of obtaining products, quotients, etc., applied to the ordinates of standard curves.

[1], [2], [3], [4] are equal uniform scales along OX, OY, OU, OV ; [5] and [6] are segmentary scales figured so that

$$x [1] : a [6] \rightarrow ax [3] ; u [3] : a [5] \rightarrow au [1].$$

$$y [2] : a [5] \rightarrow ay [4] ; v [4] : a [6] \rightarrow av [2].$$

Also, if $x [1] : u [3]$ and $y [2] : v [4]$ intersect on OO' ; then, $y/x = v/u$.

Consider any two points, $Z_1(X_1 Y_1)$ on Curve I., and $Z_2(X_2 Y_2)$ on Curve II.

To find $Y = Y_1 Y_2$. If $x = Y_1$; $v = Y_2$, $u = 1$; then $y = Y$.

To find $Y = Y_1/Y_2$. If $x = 1$; $v = Y_1$, $u = Y_2$; then $y = Y$.

To find $Y = 1/Y_1$. If $x = 1$; $v = 1$; $u = Y_1$; then $y = Y$.

5.4 illustrates the operation $Y = Y_1 Y_2$.

$$Z_1 \rightarrow A \rightarrow x [X] ; Z_2 \rightarrow B [Y] \rightarrow 1 [5] \rightarrow v [V].$$

$$x [X] : 1 [U] \rightarrow C [5] ; v [V] : C [5] \rightarrow D [Y] \quad D \text{ or } Y = 0.08.$$

If the scale of representation of Y is unsuitable, [5] or [6] may be used to provide a suitable scale factor.

Alternatively, mark a suitable scale [7] along OY.

$$Y[1] : P \text{ or } 10[5] \rightarrow 10Y[V] : 1[5] \rightarrow Y[7].$$

If a series of values of Y is required, 1[5] and P[5] may be used as centres of projection to project Y first on [V] and then on [7].

[5] and [6] may be used to carry out such operations as $Y = (aY_1)(bY_2)$. It may be desirable in this and other cases to use auxiliary scales OU', OV'.

The method described may be used to obtain $Y = e^{-X^2} \cdot \sin \phi$ from the standard curves $Y_1 = e^{-X^2}$ and $Y_2 = \sin \phi$.

11. Summary.

One standard curve representing $Y = f(X)$ may represent all curves of the same type $y = f(px)$ by a suitable transformation of scales. This transformation is a simple projection of [X] into [x] in the case of the transformation described.

The method of scale transformation gives useful results :—

1. It leads to a general solution of any quadratic or cubic equation having real coefficients.
2. It enables the evaluation of functions to be rapidly carried out by graphical methods.
3. It provides an accurate method of obtaining derived and integral curves.
4. It provides a method of studying errors and deviations which does not involve laborious computation.

The combination of various mathematical and graphical methods enables scale transformations to be carried out rapidly with any desired choice of scales. For the application of mathematics to scientific and engineering problems, mathematical methods should not be restricted to one type of development or procedure. The methods of solid geometry (analytical or drawing), projective geometry, and graphical methods, may all be used to advantage in the same problem and may provide a more practical solution than, for example, a purely analytical treatment. Similarly, the methods of analytical geometry, projection, nomography, etc., may be combined if desirable to the solution of practical problems. Frequently, a combination of methods is to be preferred to a formal solution restricted to one particular branch of mathematics. The pure mathematician and the applied mathematician often differ in opinion with regard to elegance of treatment.

Although some of the constructions involve drawing many straight lines, fewer lines are involved than in perspective drawing. The accuracy of the methods is much greater than is indicated by the drawings provided large scale drawings are carried out with an accurate straight edge (for example, a good parallel rule or T-square).

Section 8 on normal variation and random error should find application in many ways. For example, to medical diagnosis, bacteriology, experimental results, examination statistics, mass production of articles, etc.

The methods given in Section 8 are alternative to those discussed in the following :—

B.S. 600R : 1942 (British Standards Institution).

A.S.T.M. 'Manual on Presentation of Data' (American Society for Testing Materials).

"Probability Graph Paper and its Engineering Applications" by H. Rissik, Jour. Amer. Soc. Naval Eng. vol. liv. No. 1, 1942.

Discussion on the Application of Statistical Control, Jour. Inst. Elec. Eng. lxxxix. Part 1. July, 1942, and Jour. and Proc. of Inst. of Mech. Eng. Lond. cxlvii. June 1942.

XLIII. *The Warming of Walls.*

By A. F. DUFTON, M.A., D.I.C.*

[Received February 27, 1943.]

1. IN an earlier communication dealing with the influence of the fabric of a wall †, it was pointed out that, while it is well known that some rooms are not readily warmed, complete analysis of the warming of a room is not easy. A partial analysis showed, however, and experiment confirmed ‡, that a room with masonry walls is much more readily warmed if lined with wood or other insulation of small thermal capacity.

The analysis dealt with a simple case in which the rate of heating was twice that necessary to maintain the desired temperature-difference between the faces of a wall. In the present note this restriction upon the rate of heating is removed.

2. As before, a homogeneous wall (of thickness d , conductivity k , and diffusivity h), initially at uniform temperature, is considered. Heat is supplied to one surface at a constant rate (H units per unit area, per unit time) and the temperature of the second face of the wall is assumed constant.

In time t the temperature of the wall at distance x from the heated surface increases by

$$\left\{ 1 - \frac{x}{d} - \frac{8}{\pi^2} \sum_0^{\infty} (2n+1)^{-2} e^{-(2n+1)^2 \pi^2 h t / 4 d^2} \cos (2n+1) \pi x / 2d \right\} H d / k.$$

The rise T at the surface, the limit of this as x tends to zero, is

$$\left\{ 1 - \frac{8}{\pi^2} \sum_0^{\infty} (2n+1)^{-2} e^{-(2n+1)^2 \pi^2 h t / 4 d^2} \right\} H d / k.$$

It is a practical convenience that this equals

$$\sqrt{5 h t (1 + \epsilon)} H / 2 k,$$

* Communicated by the Author.

† Phil. Mag. iv. p. 888 (1927).

‡ Phil. Mag. ix. p. 1233 (1931).

where $|\epsilon| < 0.01$ if T does not exceed two-thirds of the temperature-rise which would obtain after an infinite time.

The temperature rise is independent of the thickness of the wall, therefore, and is given with considerable accuracy by the simple formula

$$T = \sqrt{5ht} H/2k.$$

3. It may be mentioned that if the rate of heating is pkT/d , p times that necessary to maintain a desired temperature difference T between the faces of the wall, and if p is not less than 1.5, the time required to warm up the surface is approximately $4d^2/5hp^2$.

For a nine-inch brick wall $4d^2/5h$ is twenty-four hours; and the warming-up time is $24/p^2$ hours.

Department of Scientific and Industrial Research,
Building Research Station,
Garston, Hertfordshire.

XLIII. On Null Geodesics and Null-Corpuscles in the Theory of Relativity.

By C. G. PENDSE, M.A., Ph.D.*

[Received March 21, 1942.]

Introduction.

THE object of this paper † is to put forth and substantiate the following thesis :—

The rule $\delta \int ds = 0$, in the established General Theory of Relativity, is inadequate to describe the path of a particle as well as a photon (alternatively a ray of light), along whose path $ds = 0$, in the space-time continuum. It is necessary to introduce a variable τ , extraneous to the continuum and related to it in much the same way as time is related to space in Newton's ideology, and the path of a particle or a photon (alternatively a ray of light) in the space-time continuum is described by the variational equation $\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0$, while $ds = 0$ along the path of a photon (alternatively a ray of light).

The paper consists of Part I. (§§ 1 and 2), Part II. (§§ 3 and 4), and Part III. (§§ 5 and 6). In Parts II. and III. use of the word "corpuscle" is made. A *particle* is a *corpuscle*, whose proper mass is different from zero; a *null-corpuscle* is a *corpuscle* whose proper mass is zero. A *null-corpuscle* would correspond to a photon (in the current phraseology).

* Communicated by the Author.

† It might appear that some of the results in the present paper are similar to or particular cases of those obtained by other writers. It is felt, however, that, in their proper context, the results are new, in spite of their apparent likeness to or dependence upon those established by other authors.

Eddington and Tolman use the expression "ray of light" instead of the word "photon" in connexion with the above-mentioned variational rule. Milne uses the expression "light pulse" in this connexion. Hence, in order to avoid a possible disputable point "(alternatively a ray of light)" is written above. In Part I. the expression "ray of light" alone is used. In Parts II. and III. the word "corpuscle" is used.

In § 1 the question of the definition of a geodesic is considered. It is shown that the definition $\delta \int ds = 0$ leads to an impasse in the case of a null line. A formal definition is adopted and a variable τ is involved.

In § 2 the question of null geodesics in relation to the General Theory of Relativity is considered, and the thesis is put forth.

In Part II the question of null-corpuscles in the Restricted Theory of Relativity is considered. In § 3 it is submitted that the commonly accepted method of deriving the properties of a light quantum (photon) by regarding it as a limiting case of a particle is open to objection. In § 4 the question is examined from a formal point of view. The existence of a null-corpuscle and the frequency property are established, and it is shown how the quantum hypothesis is suggested. The energy of a null-corpuscle is left undetermined by the Lorentz frame.

In Part III. a provisional interpretation of τ is given.

The summation convention of dummy suffixes is used.

The list of references will be found at the end.

Acknowledgments.

I wish to thank the following gentlemen :—

- (1) Professor H. F. Baker, who gave the reference to Cayley's paper.
- (2) Professor E. A. Milne.

After the thesis developed here was first arrived at, the original manuscript was sent to Professor Milne, as his monograph was the latest book in which the rule $\delta \int ds = 0$ was mentioned. [‘Remarks on general relativity,’ §§ 504–514, pp. 341–347. Particularly §§ 507 and 508.]

In his reply, Professor Milne invited me to interpret τ in terms of a space trajectory Q_0Q_1 corresponding to a null geodesic P_0P_1 , which describes the path of a photon (alternatively a light pulse) in space-time, a point P of P_0P_1 corresponding to a definite point Q of Q_0Q_1 .

He wondered if it was associated with "the rigid body length Q_0Q , i. e., with the time taken by the light pulse to go from Q_0 to Q ." And he suggested that "it would be sufficient, as an example, to consider the case in which the space-time used for P_0P_1 is pseudo-Euclidean with $ds^2 = c^2dt^2 - \sum dx^2$ *."

* In the first sentence the words between the quotation marks are a paraphrase of the relevant passage in Professor Milne's letter. In the second sentence the words between the quotation marks are taken from his letter.

Professor Milne went through my notes and explanations thereon ; and he had given them to Dr. A. G. Walker for the same purpose.

(3) Professor Sir A. S. Eddington.

I had a talk with Sir Arthur Eddington.

PART I.

§ 1. *The Definition of a Geodesic in a Riemannian Continuum.*

Let $\{x^\alpha\}$ be the co-ordinates of a point in the continuum, which will be four-dimensional in so far as the subsequent sections are concerned, and let the line element ds be given by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu,$$

where ds^2 is positive. ds is taken to be positive ; and it is assumed, without any loss of generality, that $g_{\mu\nu} = g_{\nu\mu}$. It is also assumed that the matrix $\{g_{\mu\nu}\}$ is regular, and its inverse is denoted by the matrix $\{g^{\mu\nu}\}$. The matrix $\{g^{\mu\nu}\}$ is also symmetrical. The two three-index symbols, the Riemann-Christoffel symbols, which are involved in the course of the work, are :—

$$\begin{aligned} \begin{bmatrix} \alpha & \beta \\ \lambda \end{bmatrix} &= \frac{1}{2} \left(\frac{\partial g_{\alpha\lambda}}{\partial x^\beta} + \frac{\partial g_{\beta\lambda}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\lambda} \right), \text{ and} \\ \begin{Bmatrix} \alpha & \beta \\ \mu \end{Bmatrix} &= g^{\mu\lambda} \begin{bmatrix} \alpha & \beta \\ \lambda \end{bmatrix}. \end{aligned}$$

dx^α is a perfect differential, but not an invariant. ds is not a perfect differential ; but it is an infinitesimal invariant.

Consider, now, a curve Γ . The co-ordinates of a point on Γ are functions of some parameter. If Γ is *not* a null line, this parameter can be conveniently identified with “ s ,” “the length of the curve from some prescribed point to a variable point of the curve.” If Γ is a null line, no such identification with the line-integral of an infinitesimal invariant is possible.

Let Γ' be another curve in the continuum contiguous to Γ , and let there be a one-one correspondence between points of Γ and Γ' , a point P of Γ corresponding to a point Q of Γ' .

Let $\{\xi^\mu\}$ be the co-ordinates of a point P of Γ , and

let $\{\xi^\mu + \delta\xi^\mu\}$ be the co-ordinates of the corresponding point Q of Γ' .

It is assumed that $\delta\xi^\mu$ and its derivatives are small quantities of the first order. Let P_1 and P_2 be the end points of an arc of Γ , and let Q_1 and Q_2 be the end points of the corresponding arc of Γ' .

The expression

$$\int_{\Gamma'}^2 ds - \int_{\Gamma}^2 ds$$

will now be considered ; and it will be shown that it can be expressed

in terms of the first order quantities $\delta\xi^\mu$ (so far as the variations are concerned) *alone*, only if Γ is not a null line.

In what follows, the symbol ds and other quantities on the right-hand side will refer to the curve Γ .

$$\begin{aligned} \int_{\Gamma''}^2 ds - \int_{\Gamma}^2 ds &= \int_{\Gamma}^2 \sqrt{g_{\mu\nu}(\xi + \delta\xi) d(\xi^\mu + \delta\xi^\mu) d(\xi^\nu + \delta\xi^\nu)} - \int_{\Gamma}^2 \sqrt{g_{\mu\nu}(\xi) d\xi^\mu d\xi^\nu} \\ &= \int_{\Gamma}^2 \sqrt{g_{\mu\nu}(\xi) d\xi^\mu d\xi^\nu + \frac{\partial g_{\mu\nu}(\xi)}{\partial \xi^\alpha} d\xi^\mu d\xi^\nu \delta\xi^\alpha + g_{\mu\nu}(\xi) d\xi^\mu d(\delta\xi^\nu) + g_{\mu\nu}(\xi) d(\delta\xi^\mu) d\xi^\nu} \\ &\quad - \int_{\Gamma}^2 \sqrt{g_{\mu\nu}(\xi) d\xi^\mu d\xi^\nu} \\ &= \int_{\Gamma}^2 \sqrt{ds^2 + \frac{\partial g_{\mu\nu} d\xi^\mu d\xi^\nu \delta\xi^\alpha}{\partial \xi^\alpha} + g_{\mu\nu} d\xi^\mu d(\delta\xi^\nu) + g_{\mu\nu} d(\delta\xi^\mu) d\xi^\nu} - \int_{\Gamma}^2 ds \end{aligned}$$

as far as the first order of small quantities.

If Γ is *not* a null line,

$$\begin{aligned} \int_{\Gamma''}^2 ds - \int_{\Gamma}^2 ds &= \int_{\Gamma}^2 \left[\sqrt{1 + \frac{\partial g_{\mu\nu} d\xi^\mu d\xi^\nu}{\partial \xi^\alpha} \frac{d\xi^\alpha}{ds} \delta\xi^\alpha + g_{\mu\nu} \frac{d\xi^\mu}{ds} \frac{d}{ds} (\delta\xi^\nu) + g_{\mu\nu} \frac{d}{ds} (\delta\xi^\mu) \frac{d\xi^\nu}{ds} - 1} \right] ds \\ &= \frac{1}{2} \int_{\Gamma}^2 \left[\frac{\partial g_{\mu\nu} d\xi^\mu d\xi^\nu}{\partial \xi^\alpha} \frac{d\xi^\alpha}{ds} \delta\xi^\alpha + g_{\mu\nu} \frac{d\xi^\mu}{ds} \frac{d}{ds} (\delta\xi^\nu) + g_{\mu\nu} \frac{d}{ds} (\delta\xi^\mu) \frac{d\xi^\nu}{ds} \right] ds \\ &= \left[g_{\mu\nu} \frac{d\xi^\mu}{ds} \delta\xi^\nu \right]_{\Gamma}^2 - \int_{\Gamma}^2 \left\{ g_{\mu\nu} \left(\frac{d^2 \xi^\mu}{ds^2} \right) + \left\{ \begin{matrix} \alpha & \beta \\ \mu \end{matrix} \right\} \frac{d\xi^\alpha}{ds} \frac{d\xi^\beta}{ds} \right\} ds, \end{aligned}$$

as far as the first order of small quantities, by integrating by parts.

Hence it follows that, at points along a non-null line

$$\frac{d^2 \xi^\mu}{ds^2} + \left\{ \begin{matrix} \alpha & \beta \\ \mu \end{matrix} \right\} \frac{d\xi^\alpha}{ds} \frac{d\xi^\beta}{ds}$$

is a component of a contravariant vector, and that the equations of a non-null geodesic, when it is defined by the variational equation $\delta \int ds = 0$, are

$$\frac{d^2 x^\mu}{ds^2} + \left\{ \begin{matrix} \alpha & \beta \\ \mu \end{matrix} \right\} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds} = 0, \quad \dots \dots \dots (M)$$

where

$$g_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds} = 1.$$

On the other hand, it is easy to verify independently that $g_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds} =$ constant is an integral of the equations (M). Denoting this constant by C_M^2 , where $C_M > 0$ for obvious reasons, it is seen that the domain of C_M for non-null geodesics, defined by the equation $\delta \int ds = 0$, consists of a single number, 1.

If Γ is a null line

$$\int_1^2 ds - \int_1^2 ds = \int_1^2 \sqrt{\frac{\partial g_{\mu\nu}}{\partial \xi^\alpha} d\xi^\mu d\xi^\nu \delta \xi^\alpha + g_{\mu\nu} d\xi^\mu d(\delta \xi^\nu) + g_{\mu\nu} d(\delta \xi^\mu) d\xi^\nu},$$

and it is impossible to get rid of the differentials of the variations $\delta \xi^\mu$. Consequently the definition $\delta \int ds = 0$ for a geodesic leads to an impasse in the case of a null line.

But the form of the equations (M) is such that, though the definition from which they are derived leads to an impasse in the case of a null line, they can be taken as the definition of a geodesic (null or non-null), provided that the independent variable (different from " s ," the arc-length) has a continuous range for *all* curves.

Let the equations, defining a geodesic, be

$$\frac{d^2 x^\mu}{d\tau^2} + \left\{ \begin{matrix} \alpha & \beta \\ \mu \end{matrix} \right\} \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = 0. \quad (F)$$

The nature of the parameter τ is not precisely stated at this stage.

$g_{\alpha\beta} \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = \text{constant}$ is an integral of the equations (F); let this constant be denoted by C_F^2 , where $C_F \geq 0$ for obvious reasons. Consequently, along a geodesic, $\frac{ds}{d\tau} = C_F$. For a non-null geodesic $C_F > 0$ and for a null geodesic $C_F = 0$. Hence $d\tau$ would have to be at least an infinitesimal point invariant. The results can be summarized as follows* :—

Curve.	Range of " s ."	Range of C_M .	Range of C_F .
Non-null geodesic.	Continuous range of positive numbers beginning with 0.	Isolated value 1.	The set of positive numbers.
Null geodesic.	Isolated value, which may be conveniently taken to be 0.	Non-existent.	Isolated value 0.

The range of τ is a continuous range of positive numbers.

The equations (F) are the expression of the variational equation $\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0$.

* Cayley proved (*loc. cit.* p. 162) that "circular curves," *i. e.* null lines, are geodesic curves for a quadric surface. But the result is not valid for more than two dimensions; an example is furnished by Eisenhart (*loc. cit.* pp. 50-51).

This is (most probably) the only variational form, which is sufficiently wide so as to include null as well as non-null geodesics *.

§ 2. *Null Geodesics in the General Theory of Relativity: the Parameter τ .*

It will be convenient to state very briefly the relevant portion of the established General Theory of Relativity.

(i) Space-time is a fourfold Riemannian continuum.

(ii) For the fourfold continuum in question $G_{\mu\nu}=0$ in empty space (Einstein's law of gravitation) †. $G_{\mu\nu}$ is a component of the "Einstein tensor."

$$G_{\mu\nu} = \left\{ \begin{matrix} \mu & \sigma \\ \alpha & \end{matrix} \right\} \left\{ \begin{matrix} \alpha & \nu \\ \sigma & \end{matrix} \right\} - \left\{ \begin{matrix} \mu & \nu \\ \alpha & \end{matrix} \right\} \left\{ \begin{matrix} \alpha & \sigma \\ \sigma & \end{matrix} \right\} + \frac{\partial}{\partial x^\nu} \left\{ \begin{matrix} \mu & \sigma \\ \sigma & \end{matrix} \right\} - \frac{\partial}{\partial x^\sigma} \left\{ \begin{matrix} \mu & \nu \\ \sigma & \end{matrix} \right\}.$$

(iii) The path of a particle or a ray of light in space-time as defined above is given by the variational equation $\delta \int ds = 0$, while in addition $ds=0$ along the path of a ray of light. It is usual to express this by saying that the path of

- (1) a particle is a non-null geodesic,
- (2) a ray of light is a null geodesic.

It will be most convenient now to consider the space-time whose metric is given by Schwarzschild's line element, as the first success of the theory lay in that it explained the discrepancy regarding the motion of the perihelion of Mercury and it also predicted the deflexion of light (Solar eclipse observations).

The mathematical part of the underlying theory consists in obtaining a solution of the field equations $G_{\mu\nu}=0$ with one singularity, *i. e.* in obtaining a set of formulæ for the field potentials $g_{\mu\nu}$ when there is one particle, and then in finding the equations of geodesics in such a continuum. In a solution thus obtained intuitive suggestions such as spherical symmetry play an important part.

* This point is specifically mentioned because, in the course of correspondence with Professor Milne, the question arose whether $\delta \int \left(\frac{ds}{d\tau} \right)^n d\tau = 0$, where n is a positive integer, or even $\delta \int \left(\text{any polynomial in } \frac{ds}{d\tau} \right) d\tau = 0$ might not be an alternative variational equation. It has been verified that $\delta \int \left(\frac{ds}{d\tau} \right)^n d\tau$ can be expressed in terms of the variations *alone*, as far as quantities of the first order of smallness are concerned, only if $n=2$ in the case of null lines. This can be seen even by considering $\int_1^2 ds - \int_1^2 ds$ which is developed above. Consequently,

the variational form $\delta \int \left(\text{any polynomial in } \frac{ds}{d\tau} \right) d\tau = 0$ is also unsuitable.

† It is not necessary to consider the refined form of the law of gravitation involving the cosmical constant.

Eddington states such a solution in § 38 [Schwarzschild's line element, equation (38.8)] and deduces the equations of planetary orbits in § 39 [equations (39.61) and (39.62)]. And later, in § 41, he deduces the equation of the path of a ray of light [equation (41.1)] by putting $h = \infty$ in order to make $ds = 0$.

In this connexion Tolman appears to have followed Eddington's book very closely.

But it has been shown in § 1 above that the definition $\delta \int ds = 0$ leads to an impasse when it is applied to a null line. And it would seem that Eddington and Tolman start with the equations of a non-null geodesic and obtain the equations of a null geodesic by multiplying both sides of equations by 0 or ∞ , while the very equations thus treated apparently do not exist.

However, in this case, by using the generalized definition of a geodesic, given in § 1 above, one can obtain the equations of a null geodesic; and the equations thus obtained agree with the equation of the path of a ray of light as stated by Eddington.

It is possible to argue in this connexion that a null geodesic is conceived as a limit of non-null geodesics, and that for this reason h has been made infinite in § 41 of Eddington's treatise*. It is hardly necessary to repeat that in the limiting case the basic definition breaks down.

A close comparison of § 1 above and §§ 39 and 41 of Eddington's treatise will show that making $h \rightarrow \infty$ amounts to making C_F (of § 1 above) $\rightarrow 0$.

Now, however, arises the question: *In what way is the parameter τ related to the space-time continuum?*

It was stated in § 1 above that the equations

$$\frac{d^2 x^\mu}{d\tau^2} + \left\{ \begin{matrix} \alpha & \beta \\ \mu & \end{matrix} \right\} \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = 0$$

are equivalent to the variational equation

$$\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0,$$

and that along a geodesic $\frac{ds}{d\tau} = C_F$.

The variational equation $\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0$ reduces to the form $\delta \int ds = 0$ for a non-null geodesic provided the constant $C_F (> 0)$ is kept unvaried.

But there is nothing with which $d\tau$ can be associated in the case of a null-line, except the differential of a particular member of a system of co-ordinates: there is no other infinitesimal invariant point function except ds . On the other hand, an important feature of the statement $\delta \int ds = 0$ is that ds is an infinitesimal invariant not depending on any one

* This argument was put forth by Sir Arthur Eddington in the course of conversation.

co-ordinate system*. Hence, if one wishes to preserve the above character, it would be necessary to assume that τ is extraneous to the continuum and is much the same to the space-time continuum as what time was to space in Newton's ideology †.

It may be argued that τ is merely useful for mathematical purposes, and that once it is eliminated it would cease to matter. Again we have an analogy from Newtonian Mechanics: the instance in question is the derivation of the equation of a central orbit by making use of the angular momentum integral.

The conclusion would seem to be that the rule $\delta \int ds = 0$ is inadequate to describe the path of a particle as well as a ray of light (along whose path $ds=0$) in the space-time continuum, that it is necessary to introduce a variable τ , extraneous to the continuum and related to it in much the same way as time is related to space in Newton's ideology, and that the path of a particle or a ray of light is described by the variational equation $\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0$, while $ds=0$ along the path of a ray of light ‡.

PART II.

NULL-CORPUSCLES IN THE RESTRICTED THEORY OF RELATIVITY.

In § 3 it is submitted that the commonly used method of deriving the properties of photons by regarding them as limiting cases of particles is open to objections. In § 4 the question is examined from a formal point of view; the frequency property of a null-corpuscle is isolated and it is shown that the energy of a null-corpuscle is left undetermined by the Lorentz space-time frame.

§ 3. *The Momentum and Energy of a Particle in the Restricted Theory of Relativity and Photons.*

In relativistic mechanics the equations of motion of a particle of proper mass m (>0) are

$$\frac{d}{dt} \left(\frac{m\dot{x}}{\sqrt{1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}}} \right) = - \frac{\partial V}{\partial x},$$

* This point has been stressed by Tolman (*loc. cit.* pp. 172-173). Levi-Civita apparently passes over this point at first (*loc. cit.* pp. 330-334); but he mentions it later (*loc. cit.* p. 337).

† If we were to take the alternative course of supposing that τ is a co-ordinate belonging to some co-ordinate system, the important point about the rule $\delta \int ds = 0$ that it is independent of any one system of co-ordinates would not be preserved.

‡ The form of the variational equation $\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0$ is not invariant if τ is replaced by a function of itself (excepting of course a linear function).

$$\frac{d}{dt} \left(\frac{m\dot{y}}{\sqrt{1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}}} \right) = -\frac{\partial V}{\partial y},$$

$$\frac{d}{dt} \left(\frac{m\dot{z}}{\sqrt{1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}}} \right) = -\frac{\partial V}{\partial z}.$$

where x, y, z are the position co-ordinates of the particle in the Lorentz frame and V is the potential function of the field, and in addition $\dot{x}^2 + \dot{y}^2 + \dot{z}^2 < c^2$. If p_x, p_y, p_z are the components of the momentum of the particle,

$$p_x = \frac{m\dot{x}}{\sqrt{1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}}}, \text{ etc.}$$

By the usual method we obtain the Hamiltonian function.

$$H = c\sqrt{m^2c^2 + p_x^2 + p_y^2 + p_z^2} + V,$$

and there exist the canonical equations

$$\dot{x} = \frac{\partial H}{\partial p_x}, \dot{p}_x = -\frac{\partial H}{\partial x}; \quad \dot{y} = \frac{\partial H}{\partial p_y}, \dot{p}_y = -\frac{\partial H}{\partial y}; \quad \dot{z} = \frac{\partial H}{\partial p_z}, \dot{p}_z = -\frac{\partial H}{\partial z}.$$

The energy of the particle is given by

$$E = \frac{mc^2}{\sqrt{1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}}}, \text{ and } p_x, p_y, p_z, p_t \left(= \frac{E}{c^2} \right)$$

form a four-vector for Lorentz transformations of the variables x, y, z, t .

The process by which the Hamiltonian function and the energy are obtained is untenable in the limiting case $m=0$ and $\dot{x}^2 + \dot{y}^2 + \dot{z}^2 = c^2$.

However, one comes across a limiting process, in books on Relativity or Wave Mechanics, by which the following statement is established:—

For a photon of frequency ν ,

proper mass = 0, relative mass = μ , momentum = μc , energy = $h\nu$

or energy = $h\nu$, momentum = $\frac{h\nu}{c}$.

It will be convenient to quote the relevant passage from McCrea's monograph.

McCrea (*loc. cit.* p. 39):

“ § 30. *Light Quanta*. Consider a particle having proper mass m and velocity w , then its relative mass, momentum, and kinetic energy are given by (36), (37), (41). Now let $m \rightarrow 0$ and $w \rightarrow c$ in such a way that

$$m / \sqrt{1 - \frac{w^2}{c^2}} \rightarrow \mu, \quad \dots \dots \dots (89)$$

where μ is a definite limit. Then from (36), (37), (41) we see that the particle has in the limit: proper mass zero, relative mass μ , momentum μc , energy μc^2 .”

"Now a particle moving with the velocity of light, and having a definite energy, we naturally identify with a quantum of radiation (photon). We should identify this energy with $h\nu$, where h is Planck's constant, and ν the frequency of the radiation. The quantum would then have

$$\text{Proper mass}=0, \text{relative mass } \frac{h\nu}{c^2}, \text{momentum } \frac{h\nu}{c^2}c, \text{energy } h\nu. \quad (90)$$

We need to see, however, that such an identification is consistent with the relativity principle. . . ."

The results (36), (37), (41) in the passage quoted above give the standard formulæ for relative mass, momentum, and energy.

L. de Broglie has also proceeded on these lines (*loc. cit.* Chapter X. § 1, pp. 122-123).

The following question would seem to arise (in McCrea's notation) :—

If $m \rightarrow 0$ as $w \rightarrow c$ so that $\frac{m}{\sqrt{1 - \frac{w^2}{c^2}}} \rightarrow$ a finite limit μ , what does the process

mean? Is one to assume that for every frequency ν there exists in the background an infinite hypothetical population of particles moving with constant speeds converging to c and endowed with proper masses converging to zero in such a way that

$$\frac{m}{\sqrt{1 - \frac{w^2}{c^2}}} \rightarrow \mu \left(= \frac{h\nu}{c^2} \right) \text{ as } w \rightarrow c.$$

L. de Broglie does, in fact, envisage such an infinite hypothetical population (*loc. cit.* Chapter X. § 1, pp. 122-123).

The idea of the infinite hypothetical population mentioned above appears to be in disagreement with the hypothesis that below a certain limit only a finite number of distinct values for the proper mass (associated with the fundamental "elementary particles") occur in nature (the atomic theory).

Also a corpuscle which has to have zero proper mass and which has to have the same speed c relative to all observers is something of a singular entity.

That there is a certain radical difference between photons and particles can be illustrated by the comparison of the "statistics" for photons and electrons. For photons the Einstein-Bose statistics is found to be suitable, while for electrons the Fermi-Dirac statistics is needed.

For these reasons an attempt to investigate the case of null-corpuscles is made in the following section.

§ 4. *The Hamiltonian Kinematics of a Corpuscle in the Restricted Theory of Relativity.*

It was mentioned in the preceding section that the process by which the Hamiltonian equations, the Hamiltonian function and the expression

for energy are obtained in the case of a particle cannot be used if m (proper mass) is zero and the speed is equal to c . But the final results, viz., the Hamiltonian function, the canonical equations, and the four-vector property of $p_x, p_y, p_z, p_t \left(= \frac{E}{c^2} \right)$ are such that they can be taken to be the basic hypotheses rather than derived results.

The following assumptions are made:—

(1) Associated with the motion of a corpuscle are the components of momentum p_x, p_y, p_z and energy $E = c\sqrt{m^2c^2 + p_x^2 + p_y^2 + p_z^2}$, where $m (\geq 0)$ is the proper mass of the corpuscle.

(2) $p_x, p_y, p_z, p_t \left(= \frac{E}{c^2} \right)$ form a four-vector for Lorentz transformations of the variables x, y, z, t .

(3) There exists a function V independent of p_x, p_y, p_z and E and such that, if $H = c\sqrt{m^2c^2 + p_x^2 + p_y^2 + p_z^2} + V$,

$$\dot{x} = \frac{\partial H}{\partial p_x}, \dot{p}_x = -\frac{\partial H}{\partial x}; \quad \dot{y} = \frac{\partial H}{\partial p_y}, \dot{p}_y = -\frac{\partial H}{\partial y}; \quad \dot{z} = \frac{\partial H}{\partial p_z}, \dot{p}_z = -\frac{\partial H}{\partial z}.$$

It would be useful to examine the limitations on the speed and the proper mass within which the canonical equations would be valid subject to the condition that

$x, y, z, t; \dot{x}, \dot{y}, \dot{z}; p_x, p_y, p_z$ and E be real. The kinematical equations only are involved.

We obtain the equations

$$\dot{x} = \frac{cp_x}{\sqrt{m^2c^2 + p_x^2 + p_y^2 + p_z^2}}, \quad \dot{y} = \frac{cp_y}{\sqrt{m^2c^2 + p_x^2 + p_y^2 + p_z^2}}, \quad \dot{z} = \frac{cp_z}{\sqrt{m^2c^2 + p_x^2 + p_y^2 + p_z^2}},$$

wherefrom follow the equations

$$\left. \begin{aligned} p_x^2(\dot{x}^2 - c^2) + p_y^2\dot{x}^2 + p_z^2\dot{x}^2 + m^2c^2\dot{x}^2 &= 0 \\ p_x^2\dot{y}^2 + p_y^2(\dot{y}^2 - c^2) + p_z^2\dot{y}^2 + m^2c^2\dot{y}^2 &= 0 \\ p_x^2\dot{z}^2 + p_y^2\dot{z}^2 + p_z^2(\dot{z}^2 - c^2) + m^2c^2\dot{z}^2 &= 0 \end{aligned} \right\} \quad \dots \quad (P)$$

for p_x^2, p_y^2, p_z^2 .

It is easy to verify that in order that the equations (P) be consistent and yield real values for p_x, p_y, p_z in terms of $\dot{x}, \dot{y}, \dot{z}$, it is necessary and sufficient* that

$$(a) \quad \dot{x}^2 + \dot{y}^2 + \dot{z}^2 < c^2, \quad \text{or} \quad (b) \quad \dot{x}^2 + \dot{y}^2 + \dot{z}^2 = c^2 \text{ and } m = 0.$$

$$(a) \quad p_x = \frac{m\dot{x}}{\left(1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}\right)^{\frac{1}{2}}}, \quad p_y = \frac{m\dot{y}}{\left(1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}\right)^{\frac{1}{2}}}, \quad p_z = \frac{m\dot{z}}{\left(1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}\right)^{\frac{1}{2}}}.$$

If

$$m = 0, \quad p_x = p_y = p_z = 0.$$

$$(b) \quad p_x = \frac{E}{c^2}\dot{x}, \quad p_y = \frac{E}{c^2}\dot{y}, \quad p_z = \frac{E}{c^2}\dot{z}.$$

* Cf. Bôcher, *loc. cit.* Chapter IV.

where E , the energy of the corpuscle, is arbitrary and therefore left *undetermined* by the motion (Lorentz space-time frame).

It will be desirable to make clear the status of p_x, p_y, p_z , and E as they are involved here. They appear to be conceptual. As their original definitions become meaningless when $m=0$ and $\dot{x}^2+\dot{y}^2+\dot{z}^2=c^2$, the scaffolding leading to them is discarded and the way backwards from the kinematic canonical equations is taken. The original scaffolding is recovered when $\dot{x}^2+\dot{y}^2+\dot{z}^2<c^2$. But the method leads to a region, null-corpuscles (whose proper mass is zero), in which the original scaffolding is not obtained in its own right. Finally, since the energy of a null-corpuscle moving with speed c is left undetermined by the Lorentz frame, its momentum and energy *are* conceptual to a certain extent. An infinite variety of null-corpuscles can exist. Since the momentum and energy of a null-corpuscle moving with speed less than c vanish identically, such null-corpuscles are not considered here.

Now we consider the transformation of the energy of a null-corpuscle when a Lorentz transformation is employed.

Consider the Lorentz transformation

$$x' = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} (x - vt), \quad y' = y, \quad z' = z, \quad t' = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \left(t - \frac{vx}{c^2}\right),$$

where $v^2 < c^2$.

We use primes for quantities connected with the primed co-ordinate system.

Then, since $p_x, p_y, p_z, \frac{E}{c^2}$ are assumed to form a four-vector,

$$p'_x = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \left(p_x - v \frac{E}{c^2}\right), \quad p'_y = p_y, \quad p'_z = p_z, \quad \frac{E'}{c^2} = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \left(\frac{E}{c^2} - \frac{v}{c^2} p_x\right).$$

But $p_x = \frac{E}{c^2} \dot{x}$, $p_y = \frac{E}{c^2} \dot{y}$, $p_z = \frac{E}{c^2} \dot{z}$ for any corpuscle.

Therefore,

$$p'_x = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \left(p_x - v \frac{E}{c^2}\right), \quad p'_y = p_y, \quad p'_z = p_z, \quad E' = E \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \left(1 - \frac{v}{c^2} \dot{x}\right).$$

Let ϕ be the "angle made by the direction of motion of the corpuscle with the x -axis in the unprimed system."

Then, for a null-corpuscle moving with speed c (in every Lorentz frame), $\dot{x} = c \cos \phi$ and

$$E' = E \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \left(1 - \frac{v}{c} \cos \phi\right)$$

$$\text{or} \quad \frac{E'}{E} = \left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \left(1 - \frac{v}{c} \cos \phi\right).$$

In the above formula for $\frac{E'}{E}$ the null-corpuscle is *not necessarily* moving in a straight line in the unprimed Lorentz frame.

It is important to observe that, for a Lorentz transformation, the transformation formula for the *energy* of a null-corpuscle, *moving in any*

manner in a Lorentz frame (always with constant speed c), is identical with the transformation formula for the frequency* of a plane electromagnetic wave, whose direction of propagation in the same Lorentz frame coincides with the direction of motion of the null-corpuscle at the same instant.

Hence one may be persuaded to associate a frequency, at least symbolically, with the energy of a null-corpuscle, the frequency and the direction of propagation of the associated plane electromagnetic wave being unchangeable in a definite Lorentz frame if the energy and the direction of motion of the null-corpuscle are conserved in the Lorentz frame.

On the other hand, one has to bear in mind the following points:—

(i) Unlike the energy of a particle, the energy of a null-corpuscle is left undetermined by the Lorentz frame.

(ii) Consequently there exists an infinite variety of null-corpuscles.

(iii) If by any means the speed of a null-corpuscle is reduced to less than c , the momentum and the energy of the null-corpuscle vanish identically.

The following points (iv) and (v) arise in connexion with such a frequency:

(iv) Is it to be assumed that one (and only one) frequency is associated with any null-corpuscle?

(v) Is it to be assumed that the ratio $\frac{\text{energy}}{\text{frequency}}$ is the same for all null-corpuscles (and an absolute constant) if the answer to (iv) be in the affirmative? †

(iii) is quite harmless. If the speed becomes less than c , the associated wave would cease to exist.

Considering (i) and (ii) in conjunction with (iv) and (v) an answer in the affirmative to (iv) and (v) would be the simplest answer and hypothesis.

Thus it is seen how the relation " $E = h\nu$ " for a photon is suggested, though one cannot say anything more because of (i).

But it is seen at the same time that the association of a unique frequency $\left(= \frac{E}{\text{a universal constant}} \right)$ at every point on the path ‡ of a null-corpuscle is consistent with the principle of restricted relativity §.

* Cf. Einstein, *loc. cit.* page 56: " $\nu' = \nu \frac{1 - \cos \phi \frac{v}{c}}{\sqrt{1 - \frac{v^2}{c^2}}}$."

† The ratio $\frac{\text{Energy of a null-corpuscle}}{\text{an associated frequency}}$ is an invariant in the general case.

‡ The path need not be a straight line.

§ Cf. McCrea, *loc. cit.* p. 41.

PART III.

A PROVISIONAL INTERPRETATION OF τ WHEN THE SPACE-TIME IS GALILEAN.

§ 5. *General Hypotheses concerning the Relation of τ to the Space-time Continuum and the Trajectories of Corpuscles therein.*

(1) τ is a variable extraneous to the continuum.

(2) The path of a corpuscle in the continuum is described by the variational equation

$$\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0,$$

i. e., by the equations

$$\frac{d^2 x^\mu}{d\tau^2} + \left\{ \begin{matrix} \alpha & \beta \\ \mu & \end{matrix} \right\} \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = 0,$$

and $ds^2 = 0$ along the path of a null-corpuscle.

(3) The variational equation $\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0$ reduces to the form $\delta \int ds = 0$ for a particle.

(4) At every point on the path of a corpuscle there exists a contra-variant vector $\{p^\alpha\}$, called the momentum vector, and

$$\left\{ \frac{dx^\alpha}{d\tau} \right\} = k \{p^\alpha\},$$

where k is a positive universal constant*, at every point on the path of the corpuscle.

This hypothesis is a generalization of the momentum of a particle, viz., $cm \left\{ \frac{dx^\alpha}{ds} \right\}$, where m is the proper mass of the particle, in the established theory of relativity.

There is a certain indefiniteness about k . It would also seem that there is something conceptual about the vector $\{p^\alpha\}$. This is due to the inclusion of null-corpuscles in the scheme. Such was the case in the preceding section as well.

It may be useful to indicate how the generalization is made. It was mentioned in § 1 above that the definition $\delta \int \left(\frac{ds}{d\tau} \right)^2 d\tau = 0$ reduced to the form $\delta \int ds = 0$ in the case of a non-null geodesic, provided the constant $C_F (> 0)$ of § 1 above remained unvaried for the variation. Since the proper mass of a particle remains unaltered for the variation, C_F is associated with the proper mass. Now for a particle

$$\left\{ \frac{dx^\alpha}{d\tau} \right\} = \left\{ \frac{dx^\alpha ds}{ds d\tau} \right\} = \frac{C_F}{m} \left\{ m \frac{dx^\alpha}{ds} \right\}.$$

* The choice of the symbol is, perhaps, not happy. But it is hoped that the symbol would not be attributed the standard meanings.

If this relation is to be generalized so as to extend to *all* corpuscles, the simplest hypothesis to adopt is that $\frac{C_F}{m}$ is a positive universal constant, which is denoted by ck , where $k > 0$. And then the hypothesis (4) is adopted.

§ 6. *An Interpretation for τ when the Space-time is Galilean.*

Now we consider the case in which the space-time is Galilean. $x^1 = x$, $x^2 = y$, $x^3 = z$, $x^4 = ct$; $ds^2 = c^2(dt)^2 - (dx)^2 - (dy)^2 - (dz)^2$.

We write, in the notation of four-vectors,

$$p^1 = p_x, p^2 = p_y, p^3 = p_z, p^4 = cp_t.$$

The path of a corpuscle is given by the equations

$$\frac{d^2x}{d\tau^2} = 0, \quad \frac{d^2y}{d\tau^2} = 0, \quad \frac{d^2z}{d\tau^2} = 0, \quad \frac{d^2t}{d\tau^2} = 0,$$

or
$$\frac{dp_x}{d\tau} = 0, \quad \frac{dp_y}{d\tau} = 0, \quad \frac{dp_z}{d\tau} = 0, \quad \frac{dp_t}{d\tau} = 0.$$

Let a_x, a_y, a_z, a_t denote the constant values of p_x, p_y, p_z, p_t . Then

$$x = ka_x\tau + b_x, \quad y = ka_y\tau + b_y, \quad z = ka_z\tau + b_z, \quad t = ka_t\tau + b_t,$$

where a 's and b 's are constants, and

$$(a_x, a_y, a_z, a_t) \quad \text{and} \quad (b_x, b_y, b_z, b_t)$$

are four-vectors for Lorentz transformations of the variables (x, y, z, t) ; energy of the corpuscle in the (x, y, z, t) system

$$a_t = \frac{\text{energy of the corpuscle in the } (x, y, z, t) \text{ system}}{c^2}.$$

A corpuscle moves in a straight line in the Lorentz frame (x, y, z, t) , and for a null-corpuscle $a_x^2 + a_y^2 + a_z^2 = c^2 a_t^2$.

Also for a null-corpuscle a_t will transform like the frequency* of a plane electromagnetic wave. Hence, associated with the energy of the null-corpuscle, which is constant in the Lorentz frame, there is a *unique* quantity, ka_t , which also transforms like a frequency. If, therefore, this *unique* quantity be taken to be the "*physically relevant frequency*," when once the duality contained in the association of corpuscular and undulatory representations is accepted, k can be associated with Planck's constant and c .

More generally the unique frequency associated with the null-corpuscle might have been taken to be $f(c, k)a_t$ where the function $f(c, k)$ is to be the same for all null-corpuscles. Thus†, if $f(c, k) = k$, $k = \frac{c^2}{h}$; if $f(c, k) = \frac{c^2}{k}$, $k = \hbar$, where \hbar is Planck's constant. However, the function $f(c, k)$ must not be independent of k . k is the only universal constant extraneous to the space-time continuum; if $f(c, k)$ were independent of k , the

* Cf. Part II. § 4 above.

† The standard formula " $E = \hbar\nu$ " is used here.

essential arbitrariness involved in the association of a unique frequency with a null-corpuscle would be absent. c is put in merely because it is the only other available universal constant. Hence the general remark that k is connected with c and Planck's constant.

$$\frac{x-b_x}{ka_x} = \frac{y-b_y}{ka_y} = \frac{z-b_z}{ka_z} = \tau,$$

$$i. e. \quad \frac{dx}{ka_x} = \frac{dy}{ka_y} = \frac{dz}{ka_z} = d\tau,$$

wherefrom follow the relations

$$d\tau = \frac{\sqrt{(dx)^2 + (dy)^2 + (dz)^2}}{k\sqrt{a_x^2 + a_y^2 + a_z^2}} = \frac{\sqrt{c^2(dt)^2 - (dx)^2 - (dy)^2 - (dz)^2}}{k\sqrt{c^2a_t^2 - a_x^2 - a_y^2 - a_z^2}}$$

for a particle; and

$$d\tau = \frac{\sqrt{(dx)^2 + (dy)^2 + (dz)^2}}{k\sqrt{a_x^2 + a_y^2 + a_z^2}} = \frac{\sqrt{(dx)^2 + (dy)^2 + (dz)^2}}{kca_t}$$

for a null-corpuscle, since $a_x^2 + a_y^2 + a_z^2 = c^2a_t^2$ in this case.

Denoting by E the energy of the null-corpuscle in the Lorentz frame in question, $c^2a_t = E$. E remains unchanged in the particular Lorentz frame.

$$\text{Therefore} \quad d\tau = \frac{c}{k} \frac{\sqrt{(dx)^2 + (dy)^2 + (dz)^2}}{E}$$

for a null-corpuscle; the right-hand side is invariant for Lorentz transformations. If ds_0 = the "rigid body" length element of the space-path (straight line) of the null-corpuscle,

$$d\tau = \frac{c}{k} \frac{ds_0}{E},$$

or

$$\tau = \frac{c}{k} \frac{s_0}{E} + \text{constant},$$

where s_0 is the "rigid body" length from some definite point on the space-trajectory in the Lorentz frame in question, E is the energy of the null-corpuscle and k is a universal constant connected with c and Planck's constant.

The τ "dimension" would seem to be connected in some way with energy. It is only through null geodesics in the space-time continuum that this "dimension" begins to manifest its "existence." The question awaits further investigation.

References.

- (1) A. Cayley, "On Geodesic Lines, in particular those on a Quadric Surface" (1872). 'The Collected Mathematical Papers,' vol. viii, pp. 156-180.
- (2) A. S. Eddington, 'The Mathematical Theory of Relativity' (Ed. 2, 1924 & 1930).
- (3) L. P. Eisenhart, 'Riemannian Geometry' (1926).
- (4) T. Levi-Civita, 'The Absolute Differential Calculus' (Second Impression, 1929).

- (5) R. C. Tolman, 'Relativity, Thermodynamics, and Cosmology' (1934).
- (6) M. Bôcher, 'Introduction to Higher Algebra' (1907).
- (7) L. de Broglie, 'An Introduction to the Study of Wave Mechanics' (1930).
- (8) W. H. McCrea, 'Relativity Physics' (1935).
- (9) E. A. Milne, 'Relativity, Gravitation and World-Structure' (1935).
- (10) 'The Principle of Relativity. A Collection of Original Memoirs on the Special and General Theory of Relativity.' By H. A. Lorentz, A. Einstein, H. Minkowski and H. Weyl. With notes by A. Sommerfeld. Translated by G. B. Jeffery and W. Perret (1923).

XLIV. *A Note on the Motion of a Mercury Index in a Capillary Tube.*

By H. GWYNEDD GREEN, M.A., University College, Nottingham*.

[Received November 25, 1942.]

DR. G. D. YARNOLD † has investigated the motion of mercury indices in capillary tubes and, by assuming exact observance of Poiseuille's equation for velocities below a critical one and by varying the length of the index, has deduced a value for the coefficient of viscosity. The coefficient thus obtained was about 6 per cent. too high, and in subsequent papers ‡ the same author has shown that this excess cannot be explained by electrification of the tube. The object of the present note is to present an alternative method of analysis, extending the velocity range, and to show that the results are comparable with those obtained in large-scale continuous flow experiments.

Taking the Poiseuille equation, Yarnold uses the formula

$$p = \frac{8v}{a^2}[(l+x)\eta + (L-l)\eta'] + \frac{2T}{a}(\cos \theta_1 - \cos \theta_2),$$

where p is the net driving pressure, l the length of the index with x an end correction, L the length of the tube and a its radius, η and η' the viscosities of mercury and air, T the surface tension of mercury, and θ_1 , θ_2 the angles of contact of the mercury menisci when in motion. From his results † he is able to deduce that θ_1 , θ_2 are, to a sufficient degree of accuracy, independent of the velocity. The Poiseuille law, however, dealing with an unending steady flow, has a primary assumption that there is no radial flow and no slip at the glass surface, but in the case of a finite index in which the whole of the fluid is transferred bodily, one of these conditions at least cannot hold, and we may expect some radial flow accompanied by an increase near the boundary of the axial velocity gradient.

* Communicated by Professor L. F. Bates, Ph.D., D.Sc.

† Yarnold, G. D., Proc. Phys. Soc. I. p. 540 (1938).

‡ Yarnold, G. D., Proc. Phys. Soc. lii. p. 191 (1940); and lii. p. 196 (1940).

In practical hydraulics the tractive force per square centimetre between the surface of a pipe and the liquid flowing through it is taken as being proportional to $v^2 \left(\frac{\eta}{va} \right)^n = \left(\frac{\eta}{a} \right)^s v^{2-s}$ (or the sum of such expressions), but experimental results show considerable diversions in the values of the constant of proportionality and of the power index s , variations in the circumstances of the flow causing acute changes in these values. For low velocities with stream-line flow it is generally accepted that $s=1$: for high velocities when a steady turbulent motion is established, Reynolds, for water flowing through lead pipes, obtains $n=2-s=1.70$, and Lea, for water flowing through brass pipes, obtains $n=1.72$. Between the two states there is an intermediate stage in which n reaches higher values*. If p' is the pressure necessary to drive an index of contact length l , we have, therefore,

$$\pi a^2 p' = 4\alpha v^2 \left(\frac{\eta}{va} \right)^s 2\pi a l + 2\pi a T (\cos \theta_1 - \cos \theta_2),$$

where α is some constant, or

$$p' = \frac{8}{a} \left[\alpha v^2 \left(\frac{\eta}{va} \right)^s l \right] + \frac{2T}{a} (\cos \theta_1 - \cos \theta_2).$$

For the registered pressure p we must add to p' a term Dv to allow for the effect of the air in the pipe and in the pressure cylinders. We will, then, assume empiric formulæ :

for low velocities,

$$p_1 = Av + D_1 v + B;$$

for intermediate velocities,

$$p_2 = Cv^m + D_2 v + B;$$

for higher velocities,

$$p_3 = Ev^n + D_3 v + B,$$

where, in Yarnold's experiments A , C , E should all be proportional to the length of the index, and D_1 , D_2 , D_3 should be equal and constant (the changes consequent on changes of the index length are not within the range of measurement of the apparatus). Further, taking $s=1$, we have

$$A = \frac{8}{a^2} \alpha \eta l,$$

or $\frac{Aa^2}{8l}$ is the "apparent" viscosity assuming a Poiseuille formula, and for the transitional velocities we have

$$\text{from } p_1 = p_2, \quad {}_1V_2 = (A/C)^{\frac{1}{m-1}},$$

$$\text{and from } p_2 = p_3, \quad {}_2V_3 = (E/C)^{\frac{1}{m-n}}.$$

* Lea, F. C., 'Hydraulics for Engineers and Engineering Students,' 5th edn.

An analysis of the p, v graph, assisted by the graphs of $\log z$ against $\log v$, where

$$z = \left(\frac{dp}{dv} - \frac{p-B}{v} \right),$$

which equals $(m-1)Cv^{m-1}$ or $(n-1)Ev^{n-1}$,

according as p is p_2 or p_3 , enables evaluation of all the constants with the exception of A and D_1 , of which the sum only can be obtained, so that for A the assumption $D_1=D_2=D_3$ must be used.

The range of velocities used by Dr. Yarnold in his papers is insufficient for the purposes of this analysis: he has, however, provided the author with two sets of readings of sufficient range; but, unfortunately, mercury of the highest purity was not then available, so that the figures are hardly standard. The results of the analysis of these readings are appended. The units employed for the constants are such as to give p in cm. of xylene; the units of other figures are absolute. Probable third significant figures are indicated by dropped letters; where this figure is very doubtful but the tendency known, by $+$ or $-$.

$$a=0.1014 \text{ cm.}$$

l .	$A+D_1$.	B.	C.	D_2 .	D_3 .	E.	m .	n .
20.2	.45 ₀	5.93	.027 ₅	.14 ₋	.14	.067	2.0 ₊	1.7
13.0	.35 ₀	5.93	.018	.14	.14	0.44	2.0 ₊	1.7

$m-n$.	$\alpha\eta$.	${}_1V_2$ (calc.)	(from graph)	${}_2V_3$ (calc.)	(from graph)
.31	.016 ₈	11.3	11.6	17.6	17.6
.31	.017 ₆	11.4	11.4	17.8	17.6

It will be noticed that the mean value of $\alpha\eta$ thus obtained is almost exactly that obtained by Yarnold, for this tube, by his analysis, using indices of many lengths.

While the number of readings is too few and the degree of experimental exactitude too low for absolute conclusions, the agreements among the various figures of these results may be taken to indicate that the flow of a finite index has the same characteristics as that of continuous liquid.

I have to thank Dr. Yarnold for the facilities he has allowed me in the use of his original detailed observations and for his generosity, in spite of other urgent calls, in making new observations.

XLV. *A Note on Entropy and Irreversible Processes.*

By Dr. W. EHRENBERG *.

[Received February 8, 1943.]

SUMMARY.

On the basis of an analysis of the notion "body" in thermodynamics irreversible processes are classified as "conduction" and as "breaking-up" processes. A new derivation of the relation $dS=dQ/T$ is given on the basis of Clausius's principle, and it is shown that this relation is not restricted to reversible processes, but holds for irreversible "conduction" processes as well. Bridgman's hypothesis, the theory of Carathéodory and Born and the bearing of the new derivation on the interpretation of the equilibrium conditions are discussed.

INTRODUCTION.

IN the traditional representation of thermodynamics the assertion is found that the change of entropy for a simple system is given by $dS=dQ/T$, on the assumption that no irreversible change of state has occurred during the process. This sweeping condition, sometimes expressed by $dS=dQ_{\text{reversible}}/T$, is, however, difficult to understand. The relation can, *e. g.*, be applied to the change of entropy of a system of two bodies at different temperatures that arises from a thermal conduction between them. The equalization of temperatures by heat conduction is a typical case of an irreversible process, so that here at least it is permissible to drop the restriction of reversibility. Traditionally, this case can be justified by splitting up the process of heat transfer into a great number of steps each of which consists of a transfer of heat into or from a reservoir of almost the temperature of the body. This consideration only adds complication and does not affect the result. It appears that, for a system of bodies at different temperatures, the change of entropy is not at all given by dQ/T . But the entropy of a single body is simply a function of the state of this body, and it can therefore not be of any interest how the body receives its heat, whether by a reversible or an irreversible process. A corresponding difficulty arises in the traditional discussion of the complex expressions of thermodynamics. Free energy, *e. g.*, is traditionally introduced by the expression $dS-d(U-A)/T \geq 0$. The question arises: For which case does the inequality cater? For a complex system the second term has not generally a meaning. For a simple system either the body is

* Communicated by the Author.

governed by an equation of state, when the equality holds without restriction, or the body becomes, *e. g.*, turbulent, in which case the second term loses its meaning, as a temperature can no longer be attributed to a body that has split up into parts of different temperatures.

Bridgman has made a positive case for relaxing the condition of reversibility for the expression $dS = dQ/T$. He points out that Kelvin's deduction of the thermoelectric relations suffers from the difficulty that the thermoelectric processes are regarded as reversible in spite of the fact that they are always accompanied by the irreversible processes of heat conduction and Joule heat, which by no experimental arrangement can both be reduced in relation to the processes considered. In view of these difficulties Bridgman has reconsidered the derivation of the thermoelectric relations and has shown that they are readily obtained by equating the entropy increase produced by conductive equalization of the temperature differences due to the Peltier effect and that produced by the heating up of the connecting wire as the result of the passage of the current through the potential difference existing between the junctions. "The hypothesis that the irreversible process of heat conduction and Joulean heating are always accompanied by the same characteristic increase of entropy, whether or not accompanied by other processes, must be recognized to be a new hypothesis, going beyond any explicit formulation to be found in classical thermodynamics. The new hypothesis seems to me a most natural one, certainly not opposed in spirit to classical thermodynamics, and quite consistent with statistical pictures of the nature of entropy"*.

It is the object of this paper to show that Bridgman's hypothesis is, in fact, a consequence of Clausius's principle, and to present a more general derivation of the thermodynamical relations than the traditional one.

1. Clausius has given the formulation of the second law: "Heat can never pass from a colder body to a warmer body without some other change, connected therewith, occurring at the same time," or, briefly expressed: "Heat cannot by itself pass from a colder to a warmer body" †. This statement implies that heat can pass by itself from a warmer to a colder body, and between bodies having the same temperature.

Clausius's principle is a highly condensed summary of experimental facts and theoretical considerations. Clausius himself has given detailed explanations to a number of points. As the scope of thermodynamics has rather increased since the times of Clausius, it is less astonishing that his interpretation can be supplemented in a few minor points than that

* 'The Thermodynamics of Electrical Phenomena in Metals,' New York, 1934, p. 57.

† R. Clausius, 'The Mechanical Theory of Heat,' ed. by T. Archer Hirst, 1867, London, p. 117.

he, and similarly Kelvin, succeeded in covering so wide a field by a simple formula.

A quantity of heat is defined only by reference to the first law: $Q=U+A$, where U is the increase of internal energy of a body and A the work done by the body during the process under test. The first law thus attributes a loss or gain of heat to a body for any process it undergoes if U and A are known. A is generally known from the theories of mechanics and electricity. U can be determined as a function of the variables of the body by measuring A for a process for which $Q=0$, *i. e.* for which the body is separated from other bodies by an adiabatic wall. A can partly be expressed in the external variables of the body, *e. g.* as $p.dV$. But not all work done on a body can be expressed in terms of these variables; electric work done on a body as Joule heat cannot be so expressed, nor can work done by friction*. If the work expressible in terms of the variables x_v is $X_v dx_v$, and the work not thus expressible $-B$, the first law can be written

$$dQ + dB = d(Q+B) = \frac{\partial U}{\partial t} dt + \Sigma \left(\frac{\partial U}{\partial x_v} + X_v \right) dx_v, \quad . \quad . \quad (1)$$

where t is the temperature of the body. If no work is done that cannot be expressed in terms of the variables, $B=0$; Q and $Q+B$ are completely equivalent, and any solution of $dQ=0$ is also a solution of $dQ+dB=0$. No discrimination need therefore be made between Q and $Q+B$.

Eqn. (1) actually defines a thermodynamical body by the variables in which dQ or $d(Q+B)$ can be expressed. These variables are called the

* The implications of $Q=U+A$ are :

(1) The body to which the equation applies has a finite number of observable variables, as temperature, pressure, magnetic field, etc., in which U can be expressed so that dU is a total differential in a finite number of variables of the body.

(2) It is possible to isolate the body by an adiabatic wall, *i. e.* a wall which guarantees $Q=0$, so that conditions can be created for which $U=-A$. As A is generally known from the theories of physics, U (or Δu) can be determined as a function of the variables of the body by a simultaneous observation of the variables and of A .

(3) The law of conservation of energy, excluding thermal phenomena, implies that an exchange of energy can always be observed in terms of expressions of the type $X dx$, twice, *i. e.* separately on the two bodies (or media, or fields) between which the exchange of energy takes place. If thermal phenomena are not excluded, the work done may or may not be expressible in such a way. If, *e. g.*, a piston is moved in a cylinder containing a gas, the work appears from the outside as $F.dx$, from the gas as $p.dV$, so that $F.dx = -p.dV$. But if a body is heated by Joule heat, the work appears from the outside as $V.dq$ (V potential difference, q charge), but from the body only as $i^2 r d\tau$ (i current, r resistance, τ time), which cannot be expressed in the external variables of the body, *i. e.* in variables other than t . A in the equation $Q=U+A$ refers to the work observed from the outside of the body, and includes both the work expressible in the variables of the body in terms of $X.dx$ and that work that cannot be so expressed.

variables of state, such as p , V , t , and they co-operate with the parameters on which the equation of state depends. In view of the variables and the parameters, a body must be homogeneous, *i. e.* within the boundaries of the body there must be everywhere the same temperature, density, specific heat, etc., if these quantities enter explicitly the equations. Any particular statement involving a thermodynamical theorem can be expected, without special examination, to hold only as far as its object satisfies these conditions. A quantity of gas, for instance, is a body when it is considered as homogeneous. A real or a virtual surface dividing this quantity makes it two bodies. But if it is considered as consisting of groups of molecules having different velocities, neither of these groups is a body, for no temperature can be ascribed to them. There are also processes in which a body splits up in parts of different values of the variables, connected or not connected with external work, in which for this reason the body also loses its property as a thermodynamical body. According to the variables and the parameters, a body can be a reservoir, a piece of matter, an electrical coil, a phase, or any other simple object of thermodynamical interest.

The simplest system in which a passage of heat can occur consists of two bodies with a common boundary through which the heat can pass. A passage of heat can be inferred from a change of Q of both bodies. The first law demands in this case that the changes of Q are numerically equal and opposite in sign, and Clausius's principle states that a process can only entail a loss of heat by the warmer body and a gain by the colder body. Passage of heat refers to the net result of the process only. In the case of transfer of heat by radiation, transfer of heat takes place in both directions, but the net result is always an equalization of temperatures.

The essential meaning of passage of heat can only, however, be appreciated with reference to a more complex system. Let this system consist of the elements a, b, \dots, l, m, n . These elements may partly be bodies in the thermodynamical sense and partly mechanical and electrical contrivances. Let us further consider a particular state A of the system, defined by values of all the parameters of the system and a process by which this state A is changed into a state B . In particular, the pair of bodies l, m may be singled out. Then the process may involve these bodies only, when the preceding consideration holds. In the general case, the bodies l, m are involved together with other elements, when the process is not "by itself" with regard to l and m . In this case Clausius's principle implies that whatever the process AB , it can always be made "by itself," by supplementing it by a reversible process BA' , so that the state A differs from the state A' only with respect to the bodies l and m , while the bodies l, m may have changed with respect to t and the parameters but not with respect to the differentials co-operating with the X_v in the expression for dQ . The "reservoirs" of the traditional thermodynamics automatically satisfy this condition, as they have no variables denoting work done. In other words, the process BA' has

not only to restore the states of the other bodies completely, but has also to restore the mechanical variables of the two bodies under test if they need restoring. Clausius's principle asserts that if in the state A the body l has been warmer than the body m , l cannot have gained heat during the transition ABA' and m cannot have lost heat during the same process.

The idea of this supplementary process is, of course, the idea of the Carnot cycle. The Carnot cycle shows that it is always possible, by means of a reversible engine and a standard substance, to produce mechanical energy or its equivalent by decreasing the temperature difference between two bodies or to dispose of any surplus mechanical energy by adding heat to one body and decreasing the heat content of another body. All thermodynamical cycles have substantially the function of restoring conditions to which the principle of heat transfer can be applied directly. A modified Carnot cycle of Clausius, *e. g.*, permits the addition of heat to a body by suitably directing a transfer of heat between two bodies at lower temperatures. By means of these cycles the states of all but two bodies, and of all mechanical variables of these two bodies, and of all mechanical devices of the system can be restored. For two bodies are needed to supply and to take up any surplus heat; and to these two bodies of the system the formula of Clausius does really apply. The cycles serve thus to interpret the meaning of "by itself"; the notion of entropy does not enter these considerations at all. In fact, Clausius had cleared up the problems involved in the Carnot cycle long before he introduced the entropy function. By means of these cycles all thermodynamical processes can be tested with respect to their compatibility with the second law.

The fact that a Carnot cycle is a reversible process does not require any special explanation. For all processes dealt with in other parts of physics are reversible in the same sense, *i. e.* they can be described by differential expressions so that they can take place independent of a change of all signs of the differentials. *It is only through Clausius's principle that the idea of irreversibility is introduced*, in particular that those processes are shown to be irreversible which could be reversed only by a passage of heat from a colder to a warmer body. Clausius himself has now drawn attention to two distinct types of irreversible process. An example of the first "is afforded by a gas which expands under such conditions that it has not to overcome an opposing pressure equal to its own expansive force." "The excess of force may then give rise to motions of considerable velocity in parts of the body under consideration, and these motions may subsequently be changed into the molecular motions which we call heat" *. In these processes the body is broken up in an uncontrollable manner into parts of different temperatures and different values of the other variables. These processes may or may not be connected with external work, the occurrence of

* Clausius, *l. c.* pp. 223 and 244.

breaking up is, in fact, independent of dQ as well as of A . Such a process cannot be described in terms of thermodynamical variables, and differential expressions involving thermodynamical variables do not apply to these breaking-up processes. The second type is exemplified by "the passage of heat between two bodies of different temperature" *. Joule heat is another example of the second type of irreversible process. It is now important to realize that it is this type that has been made by Clausius the basis of his principle, and that processes of this type do not entail the breakdown of the notion "body." Let, e. g., two blocks of material at different temperatures be connected by a thin, thermally conducting wire. By making the wire sufficiently thin the temperatures of the two blocks can be kept uniform throughout their volumes to any desired degree. The blocks remain, therefore, bodies during the process of thermal equalization. With regard to the connecting wire, it can either be regarded as so small that its content of all thermodynamical quantities can be neglected at any instant, or it can be regarded as an ordered assembly of small elements each of which is sufficiently homogeneous to be a body in the thermodynamical sense. All expressions involving thermodynamical variables can therefore be written so as to include irreversible processes of this "conduction" type.

Irreversible conduction processes may be said to be irreversible in their own right. Breaking-up processes are not at all really processes within the framework of thermodynamics, and are only recognized as irreversible because the result of such a process can be shown by means of a Carnot cycle to be equivalent to a transfer of heat from a warmer to a colder body. *It appears, therefore, advisable to classify irreversible processes into conduction and breaking-up processes.* It will be shown in the following section that the relation $dS=dQ/T$ or $dS=d(Q+B)/T$ can be derived without basing its introduction on reversible processes in particular. The conceptions "body" and "temperature" must of course be used. The relation holds, therefore, for reversible processes and for irreversible conduction processes, but not for breaking-up processes.

2. Let us consider a closed system (no exchange of energy except between the bodies of the system) of n bodies and a particular state A of this system that in general will be determined by at least $2n$ variables. In order to reach a second state B the process necessary will either result in

- (a) a transmission of heat from a warmer to a colder body, or
- (b) a transmission of heat from a colder to a warmer body, or
- (c) a transmission of heat between bodies of the same temperature or no transmissions at all.

All three cases refer to a process as supplemented by a Carnot cycle,

* Clausius, *l. c.* p. 244. Clausius has not, however, commented upon the distinction of the two types with respect to the preservation of the thermodynamical variables.

as explained before, so that the effect of the process is measured as the residual effect on two selected bodies of the system.

The effect of Clausius's on this classification is now, first, that it describes (a) and (c) as possible, (b) as forbidden; secondly, that it introduces in (a) an irreversible process; thirdly, that it makes this classification unambiguous. With respect to any state A, any other state B belongs to one of these classes, and to one only. Further, if a state C belongs to the same class with respect to B as B belongs with respect to A, the state C belongs also to the same class with respect to A. These relations are those that constitute an ordered sequence of numbers. For let the state A be given a particular number and the process (a) be represented by $>$, the process (b) by $<$ (or *vice versa*), the process (c) by $=$, then each state of the system can be given a number, so that the sequence of the numbers belonging to the states is unambiguously fixed. Any monotonic function of the original sequence represents the states likewise. Or, in other words, *Clausius's principle establishes a function of the states of any thermodynamical system.*

In order to see that the principle indeed makes the classification unambiguous it is only necessary to consider the classes in turn. For, if a state B would have to be reached by a process (b), this state is not accessible. It cannot, alternately, be reached by a process (a) or (c), for then heat could by itself be transported from a colder to a warmer body. The processes (a) and (c) also exclude each other. For if B is reached by a process (c), A can also be reached by a process (c) starting from a state B. But if B could also be reached by a process (a), the colder body would have got warmer and the warmer body colder, so that then A could only be reached again by a forbidden process (b). Similar considerations show that the order of states is maintained when a state C is introduced. Clausius's principle thus introduces an unambiguous order in the system through the transitions between the states, which can be expressed as a single-valued function, so that the question whether a state B can be reached from a state A is answered by the relative values of the function. For a particular choice of the function it is the entropy S of the system; generally it will be called F in the following discussion.

F must now be expressed as a function of the variables of the bodies. It is seen at once that the value of F is not altered by any adiabatic changes of the separate bodies, as states which are reached without thermal exchange must be given the same number. Considering a particular ν body of the system, F does not, therefore, vary if its variables are altered in such a way that $dQ_\nu=0$. As the changes in the variables permitted are the same for $dQ_\nu=0$ and for $d(Q_\nu+B_\nu)=0$, and as F is a function of the variables, F remains unchanged for changes of the body for which $d(Q_\nu+B_\nu)=0$. F does not depend then explicitly on all the variables of the system, but only on the parameters σ_ν which number the particular adiabatic curves or surfaces to which the states of the bodies belong:

$$F=F(\sigma_1, \sigma_2, \dots \sigma_n);$$

$$dF = \frac{\partial F}{\partial \sigma_1} d\sigma_1 + \frac{\partial F}{\partial \sigma_2} d\sigma_2 + \dots + \frac{\partial F}{\partial \sigma_n} d\sigma_n. \quad (2)$$

For a single body and an adiabatic process

$$0 = dQ_v = d(U_v + B_v) = dU_v + X_{v_1} dx_{v_1} + X_{v_2} dx_{v_2} + \dots,$$

where U_v is the internal energy and the $X_v dx_v$ denote the work done. It will be shown later that dQ_v has always an integrating factor N_v , so that no restriction of generality is introduced by considering the particular case of a fluid with the two independent variables t and V , for which the existence of an integrating factor is obvious. Let then

$$N_v dQ_v = N_v (dU_v + p_v dV_v) = d\sigma_v, \quad (3)$$

where σ_v is a function of the variables which is constant for any adiabatic process and N_v depends only on the variables of the body v .

Combining (2) and (3),

$$dF = \frac{\partial F(\sigma_1, \dots, \sigma_n)}{\partial \sigma_1} N_1(t, V_1) dQ_1 + \dots + \frac{\partial F}{\partial \sigma_n} N_n(t, V_n) dQ_n. \quad (4)$$

Let us now consider a particular process by which an isothermal exchange takes place between two bodies of a system and no exchange affecting the other bodies :

$$\begin{aligned} dF &= 0, \\ dQ_1 &= -dQ_2; \quad dQ_3 = \dots = dQ_n = 0, \\ t_1 &= t_2 = t. \end{aligned}$$

Equation (4) then becomes

$$\frac{\partial F(\sigma_1, \sigma_2, \dots, \sigma_n)}{\partial \sigma_1} N_1(t, V_1) = \frac{\partial F(\sigma_1, \sigma_2, \dots, \sigma_n)}{\partial \sigma_2} N_2(t, V_2). \quad (5)$$

As the factors on both sides of the equation depend on different variables in pairs,

$$\begin{aligned} \frac{\partial F}{\partial \sigma_1} &= \frac{\partial F}{\partial \sigma_2}, \\ N_1(t, V_1) &= N_2(t, V_2). \end{aligned} \quad (6)$$

Equations (6) can only be satisfied for general values of the variables if

$$\begin{aligned} \frac{\partial F}{\partial \sigma_1} &= \frac{\partial F}{\partial \sigma_2} = \frac{dF}{d\Sigma\sigma_v}; \quad F = \phi(\sigma_1 + \sigma_2 + \dots + \sigma_n) = \phi(s), \\ N_1(t, V_1) &= N_2(t, V_2) = N(t). \end{aligned} \quad (6a)$$

Equation (4) can thus be written

$$dF = \frac{dF}{d\Sigma\sigma_v} (N(t_1) dQ_1 + N(t_2) dQ_2 + \dots + N(t_n) dQ_n).$$

Introducing $T = 1/N(t)$ as the thermodynamical temperature scale,

$$\begin{aligned} d(\sigma_1 + \sigma_2 + \dots + \sigma_n) &= \frac{dQ_1}{T_1} + \frac{dQ_2}{T_2} + \dots + \frac{dQ_n}{T_n}, \\ d\sigma_v &= \frac{dQ_v}{T_v}. \end{aligned} \quad (7)$$

$$F = \phi \left(\int \frac{dQ_1}{T_1} + \int \frac{dQ_2}{T_2} + \dots + \int \frac{dQ_n}{T_n} \right), \dots \dots \dots (8)$$

where ϕ is an arbitrary monotonic single-valued function of one variable.

The integrating factor N for the change of heat dQ of a single body is determined by eqn. (6 a), or rather the particular factor that is compatible with eqn. (2), and, therefore, of exclusive interest. The thermodynamical scale is easily connected with the ideal gas scale, in the traditional manner. Eqn. (3) in the variables t and V becomes

$$d\sigma = \frac{1}{T} \frac{\partial U}{\partial t} dt + \frac{1}{T} \left(\frac{\partial U}{\partial V} + p \right) dV. \dots \dots \dots (9)$$

As $d\sigma$ is a perfect differential,

$$\begin{aligned} \frac{\partial^2 U}{\partial V \partial t} \cdot \frac{1}{T} &= \frac{\partial \left(\frac{1}{T} \right)}{\partial t} \left(\frac{\partial U}{\partial V} + p \right) + \frac{1}{T} \frac{\partial^2 U}{\partial V \partial t} + \frac{1}{T} \frac{\partial p}{\partial t}, \\ 0 &= \frac{\partial \left(\frac{1}{T} \right)}{\partial t} \left(\frac{\partial U}{\partial V} + p \right) + \frac{1}{T} \frac{\partial p}{\partial t}. \dots \dots \dots (10) \end{aligned}$$

For an ideal gas in particular

$$\frac{\partial U}{\partial V} = 0; \quad p = \frac{Rt}{V},$$

$$\begin{aligned} 0 &= \frac{d \left(\frac{1}{T} \right)}{dt} \cdot \frac{Rt}{V} + \frac{1}{T} \frac{R}{V} = - \frac{1}{T^2} \frac{dT}{dt} \cdot \frac{Rt}{V} + \frac{1}{T} \frac{R}{V}, \\ \frac{t}{T} \cdot \frac{dT}{dt} &= 1; \quad T = \alpha t. \dots \dots \dots (11) \end{aligned}$$

The thermodynamical scale is identical with the ideal gas scale, with an arbitrary factor.

The function ϕ of eqn. (6 a) and (8) can be chosen differently for different systems. But as separated, isolated systems cannot be compared with regard to their processes, no physical significance can be attached to this choice. Bodies that can interact must be combined in one system, when again the choice of the function has no thermodynamical significance. The statistical theory suggests $\phi = e^S$. Thermodynamically, however, the function is chosen $\phi = +S$, when ϕ is called the entropy S .

In the derivation of the expression for the entropy the system was assumed to consist of at least two bodies. A single body can undergo only adiabatic changes while the discussion leading to the proof of the existence of S is based on exchanges of heat. It is perhaps unfortunate that the name entropy has been given both to $S = \sum \int dQ_v / T_v$, where \sum refers to the different bodies and the integral to the processes of each particular body, and to $\sigma_v = \int dQ_v / T_v$. It is obvious from the derivation

that no condition with regard to reversibility was assumed for dQ/T . It is, in fact, quite irrelevant how a body receives its heat, reversibly, i. e. from a body having the same temperature, or irreversibly, i. e. from a body of higher temperature, or by Joule's heat. The only condition is that it remains a body during the process, i. e. that it does not break up. The bearing of the second law on the thermodynamics of single bodies has nothing to do with reversibility or irreversibility; it is fully contained in the statement that $1/T$ is an integrating factor for dQ . But in order to establish this factor, reference is made to the principle of irreversibility.

With the choice of $\phi(S)=+S$ the sign of the change of S for an irreversible process is also fixed. Let heat pass between two bodies without involving mechanical work. Then $dQ_2=-dQ_1$, $T_1>T_2$, dS is positive, and the second law can be restated, that for any process the entropy of a system of two bodies or more increases if it does not remain constant.

Consider now a process as a result of which a warmer body (1) has lost heat and a colder body (2) has received heat, while no other elements of the systems have gained or lost heat. Then

$$\Delta S = \frac{Q_1}{T_1} + \frac{Q_2}{T_2}, \quad \Delta S \geq 0. \quad \dots \dots \dots (12)$$

It is assumed that the bodies are so large or the quantities of heat involved so small that T_1 and T_2 can be regarded as constants. Let

$$Q = -(Q_1 + Q_2)$$

be the quantity of heat "lost," i. e. transformed into mechanical work, according to the first law, then

$$\frac{Q_1}{T_1} - \frac{Q_1 + Q}{T_2} \geq 0, \quad Q \leq -Q_1 \frac{T_1 - T_2}{T_1}.$$

The maximum amount of heat lost is therefore given by

$$-Q_1 \frac{T_1 - T_2}{T_1}, \quad \dots \dots \dots (13)$$

and the maximum mechanical efficiency of a heat engine working on the principle of the transfer of heat between bodies is given by

$$\eta = \frac{T_1 - T_2}{T_1}.$$

3. *Equilibria.*—The entropy of an isolated system can only increase or remain constant. The system is, therefore, in an equilibrium if S decreases for all virtual changes, i. e.,

$$\delta S = 0. \quad \dots \dots \dots (14)$$

The fact that the system is isolated can be expressed by the conditions

$$\left. \begin{aligned} \Sigma p \delta V &= 0 \text{ (no external work),} \\ \Sigma \delta Q &= 0 \text{ (no external heat exchange),} \\ \Sigma \delta m &= 0 \text{ (conservation of mass).} \end{aligned} \right\} \dots \dots \dots (15)$$

For a non-isolated system of n bodies, a system of $n+1$ bodies must be examined so that this system of $n+1$ bodies can be regarded as isolated. The $n+1$ th body may be considered as the external medium. Then

$$dS_{n+1} = \frac{dQ_1}{T_1} + \frac{dQ_2}{T_2} + \dots + \frac{dQ_n}{T_n} + \frac{dQ_{n+1}}{T_{n+1}} = dS + \frac{dU_{n+1} + p_{n+1}dV_{n+1}}{T_{n+1}} \quad (16)$$

can only be ≥ 0 , while

$$\left. \begin{aligned} \Sigma p_v dV_v &= 0, \\ \Sigma dQ_v &= 0, \\ \Sigma dU_v &= 0. \end{aligned} \right\} \dots \dots \dots (15)$$

Therefore

$$\begin{aligned} dU_{n+1} &= - \sum_1^n dU_v, \\ p_{n+1} dV_{n+1} &= - \sum_1^n p_v dV_v = A; \text{ A, work done on the system,} \\ T_{n+1} dS - \Sigma dU + A &\geq 0, \\ d\Sigma U - T_{n+1} dS &\leq A. \end{aligned} \quad (17)$$

T_{n+1} is then the external temperature and A the work done from outside on the system under test. If in particular the system is in thermal equilibrium in itself and with the medium

$$\begin{aligned} T_1 = T_2 = \dots = T_n = T_{n+1} &= T = \text{const.}, \\ T_{n+1} dS &= d(TS). \end{aligned}$$

Under these conditions eqn. (17) becomes

$$d(\Sigma U - TS) = A, \quad (18)$$

as no irreversible processes can take place (differentially) in a system in temperature equilibrium with an outside medium. Introducing

$$F = \Sigma U - TS$$

as the free energy of the system, eqn. (18) becomes

$$dF = A \quad (19)$$

for a system in equilibrium with a medium of constant temperature. An inequality does not arise. Similar considerations apply to the other thermodynamical functions. The main point is that relation (19) is an equation and not an inequality.

As the variables disappear for breaking-up phenomena, no differential expressions can apply to these phenomena. It is another question if any theorems can be applied to the results of breaking-up processes. There can be no doubt that Clausius meant to include these breaking-up processes in his principle. For the processes considered by Clausius, processes involved in heat engine construction, this is perfectly well founded. It will not be discussed here if it can rightly be extended to processes involving substantially statistical fluctuations. A state (2) for which the entropy is smaller than for a state (1) could always, in a

suitably enlarged system, be reached by a transfer of heat from a colder to a warmer body starting from the state (1). As this transition is impossible, S_2 must always be larger, or equal to S_1 , even if S cannot be defined for intermediate stages of the process. But if a breaking-up process is involved, $S_2 - S_1$ cannot be written as $\int dS$, the integration taken over the process, and, therefore, for a single body, not as $\int dQ/T$. In the same way it can be seen that $F_2 - F_1$ is always smaller or equal to the work done during the process if $T_1 = T_2$, even if $F_2 - F_1$ cannot be expressed as $\int dF$ and no differential expressions hold relating to F .

4. It remains to show that for a thermodynamical body dQ has always an integrating factor. For if the state of the body depends on more than two variables, $dQ = dU + A$, but both U and A must be expressed in terms of all the variables, and the existence of an integrating factor is no longer obvious. Carathéodory* has, however, shown that the existence of an integrating factor can be proved mathematically, if the assumption is made that in the neighbourhood of every state of this body there exist states which cannot be reached by an adiabatic process, *i. e.* on a solution of the equation $dQ = 0$. This assumption can be shown to follow from Clausius's principle. Carathéodory has further suggested, using a wider notion of an adiabatic process than is traditionally accepted, replacing Clausius's principle by this theorem, and basing on it all considerations concerning the second law. Carathéodory's theory has been further developed by Born†.

Let a body be changed, by adding a quantity of heat Q from a state (1) to a state (2) while keeping the mechanical variables constant, so that as the result of the process no mechanical work is done by the body. If this state (2) could also be reached adiabatically, the state (1) could also be regained by external work only, and the quantity of heat be completely transformed into work. This is impossible according to Clausius's principle, as the mechanical work could in turn be used to heat up a warmer body.

The proof of the existence of the integrating factor for dQ is based on the geometrical conception that it can be interpreted as the existence of a surface through any point of the co-ordinate space containing all curves which satisfy the equation $dQ = 0$. Let

$$dQ = X dx + Y dy + Z dz$$

and P' be a point with the co-ordinates x', y', z' . Let any surface $\psi(x, y, z) = 0$ be laid through the point P' and through the curve A consisting of the points reached by feeding Q into the body while preventing any work being done. The combination of $\psi = 0$ and $dQ = 0$ results in a two-dimensional Pfaffian for which an integrating factor exists, *i. e.* on the surface $\psi = 0$ there is one, and only one, curve going

* *Mathematische Annalen*, lxvii. p. 355 (1909).

† *Phys. Z.* xxii. pp. 218, 249, 282 (1921).

through the point P' which is a solution. Going from P' along this curve any distance l a point P is reached. If, then, the surface $\psi=0$ is deformed so that it still contains P' and P , and also the curve A , the curve on the deformed surface which represents the solution and goes through P must still go through P' . Otherwise it would cross the curve A at a point different from $Q=0$, and an inaccessible point could be reached from P' via P , which is impossible. By deforming the surface $\psi=0$ so that it always contains the points P' and P and the curve A , the curve connecting the points P' and P describes a surface containing all solutions of $dQ=0$. dQ has, therefore, an integrating factor. The proof can be extended to more than three independent variables.

Carathéodory—Born's theory thus clears the problem of the existence of the integrating factor for a thermodynamical body in general. These authors have further shown, on the basis of this theorem alone, without additional use of Clausius's or Kelvin's principle, and without reference to a cyclic process, that $1/T$ is an integrating factor for dQ , if dQ belongs to a system of bodies which have all the same temperature, or to a body belonging to such a system. Carathéodory's principle predicates, however, less than that of Clausius*, and, as a result, entropy as introduced by Carathéodory and Born cannot easily be related to irreversible processes.

The authors have, however, attempted to do so, and in order to show that S can only increase or remain constant, they discuss the adiabatic changes of a system of two bodies in temperature equilibrium. Their argument runs as follows:—Let the initial values of this system be V_1^0, V_2^0, S^0 , the final values V_1, V_2, S . Vary first the volumes adiabatically and quasistatically (*i. e.* so that the equation of state always applies) from V_1^0, V_2^0 , to V_1, V_2 . The entropy remains S^0 . Then keep V_1, V_2 constant and vary the state by adiabatic work (friction, stirring, etc.) until the entropy has changed from S^0 to S . If, now, S for different processes is sometimes larger, sometimes smaller, than S^0 , any neighbouring state could be reached adiabatically, for the volumes can be altered at liberty. This contradicts the experimental basis of the second law, or the principle in its general form. Therefore S must either be $\geq S^0$ or else $\leq S^0$ throughout.

The argument is thus based on an unusual meaning of adiabatic, on the idea that it is permissible to speak of an adiabatic process generally as of a process for which $dQ=0$, whatever else may happen to the body, so that Carathéodory's principle includes the general meaning as well as the traditional one. But the processes of stirring and friction are very ill-defined, and all that can generally be stated about them is that they are equivalent to a positive dQ . The use of adiabatic in the wider sense implies then that there are processes $dQ=0$ which are equivalent

* Ehrenfes -Afanassjeva, *Z. Phys.* xxxiii, p. 933 (1925). Planck, *S. d. Pr. Acad. d. Wiss.* p. 453 (1926).

p th iterate of x_m by $x_m^{(p)}$, and starting from zeroth values $x_1^{(0)}, x_2^{(0)}, \dots, x_r^{(0)}$, which are entirely arbitrary, we have

$$x_1^{(1)} = K_1 - \{b_1 x_2^{(0)} + c_1 x_3^{(0)} + \dots + r_1 x_r^{(0)}\}, \quad \dots \quad (2.1)$$

$$x_2^{(1)} = K_2 - \{a_2 x_1^{(1)} + c_2 x_3^{(0)} + \dots + r_2 x_r^{(0)}\}, \quad \dots \quad (2.2)$$

etc., the general expression being

$$x_m^{(p)} = K_m - \{a_m x_1^{(p)} + b_m x_2^{(p)} + \dots + l_m x_l^{(p)} + n_m x_n^{(p-1)} + \dots + r_m x_r^{(p-1)}\}. \quad \dots \quad (2.m)$$

It is convenient to put $x_1^{(0)} = x_2^{(0)} = x_3^{(0)} = \dots = x_r^{(0)} = 0$. These iterates will ultimately converge on the true values, oscillate about the true values, or diverge, according as the numerically greatest root $y = Y$ of

$$\begin{vmatrix} y & b_1 & c_1 & \dots & r_1 \\ a_2 y & y & c_2 & \dots & r_2 \\ a_3 y & b_3 y & y & \dots & r_3 \\ \dots & \dots & \dots & \dots & \dots \\ a_r y & b_r y & c_r y & \dots & y \end{vmatrix} = 0 \quad \dots \quad (3)$$

is greater than, equal to, or less than unity respectively. (Should Y be one of a pair of complex roots, the same limits apply to its modulus.) The smaller the value of Y , the more rapid, generally speaking, is the convergence, although the rate of convergence is also affected by the other roots.

Schmidt * gives an accurate method of calculating the solutions of (1) from the first r iterates of each unknown and the coefficients of (3) when expanded. It is immaterial whether the iterates converge or not, but the process takes almost as long as a determinantal solution. As a labour-saving variation, approximate solutions are obtainable by the ingenious use of the first few iterates and their differences, having first arranged the equations to give reasonably rapid convergence; this may always be done by choosing a suitable order and, if necessary, adding or subtracting to form subsidiary equations. The general idea is to make a_m, b_m , etc., in (1) appreciably less than unity, although an occasional slightly higher value does not necessarily prevent convergence. It is essential in a difference method that all the iterates used shall be free from error; consequently $(N+2)$ iterates, each taken to $(N+1)$ places, must be used to obtain results to N places, the last figure even then being slightly doubtful.

Suggested Alternative Process.

Use may be made of the fact that the number of iterations necessary to get within 10 per cent. or so of the results is commonly quite moderate, and is, moreover, reasonably independent of the values of K_1, K_2 , etc. It is also clear that, since convergence is not dependent on the initial

* Phil. Mag. p. 369 (Nov. 1941).

values $x_1^{(0)}$, $x_2^{(0)}$, etc., any errors made in the course of a convergent iteration will be automatically eliminated by subsequent iterations, provided only that the figures in the final sequence are correct. Each error, as it were, starts a new iterative system which is bound to give the same final result. It is therefore possible to work in very approximate terms in the early stages, and the method requires no checking, the penalty of an error being merely an extension in the number of iterations necessary before a steady state is reached.

It is convenient to proceed by forming a simplified set of equations from (1) (which is supposed to be already arranged in a convergent order), in which each coefficient a_m , b_m , etc. is expressed to *one decimal place only*, and the free constants K_1 , K_2 , etc. are divided by 10^n , where the integer n is selected to make the greatest quotient lie between 1 and 10; the quotients are then expressed to the first decimal place. This will be styled the first "iterative framework"; the simplified numerics will be denoted by a'_m , b'_m , K'_m , etc., and the new unknowns by X_1 , X_2 , etc. We proceed to resolve these unknowns and ultimately reconvert to the originals by using the relation $x_m = 10^n X_m$.

The first iterative framework is used to iterate all the (new) unknowns to the *units figure only*. These results are then substituted in the *original* equations (1) and, working to *two* places, the remainders R_1 , R_2 , etc. found. These remainders, divided by 10^{n-1} , form the free constants K_1'' , K_2'' , etc. of a second iterative framework, which differs from the first only in respect of the K terms, and which yields as *unit* solutions the *first decimal place* figures for the unknowns and corrects at the same time any defect in the first series of iterations, whether casual or systematic. The combined results, which now express the unknowns to the first place, are substituted in (1), working this time to *three* places, and the new remainders R_1' , R_2' , etc. are used similarly in a third iterative framework to yield the *second decimal place* for the unknowns. Proceeding in this way, the values of the unknowns are readily and systematically determined to any desired number of places, or until the remainders vanish, denoting that the solutions are exact.

Although it might be supposed that these successive series of iterations would be laborious, in actual practice they are very quickly carried out without mechanical aid, since only one figure is required in each case. In taking out the remainders, a calculating machine, although not essential, is an advantage if many significant figures are needed. A last series of remainders should be taken as a final check.

If a calculating machine is not available, the intermediate arithmetic, as far as the remainders are concerned, may be shortened by taking

$$K_m''' = 10\{K_m'' - (a'_m X_1' + b'_m X_2' + \dots + r'_m X_r')\}$$

and

$$K_m'''' = 10\{K_m''' - (a'_m X_1'' + b'_m X_2'' + \dots + r'_m X_r'')\},$$

where X_1' , X_2' , etc. are the "first place" addenda resulting from the second series of iterations, X_1'' , X_2'' , etc. the "second place" addenda from the third series of iterations, and so on. This reduces the labour.

but sacrifices the check at each stage, which is a feature of the first method. The latter is therefore to be preferred as a general rule.

Numerical Example.

Consider the equations

$5x_1+6x_2+5x_3=400,$ (4.1)

$3x_1+5x_2+4x_3=350,$ (4.2)

$x_1+2x_2+3x_3= 80,$ (4.3)

for which $Y \simeq 0.78$, and which are exactly satisfied by $x_1=-8$, $x_2=110$, $x_3=-44$ Taken as they stand, some ten iterations are required for each place, but by subtracting (4.2) from (4.1) to get (5.1) and (4.3) from (4.2) to get (5.2) we get

$2x_1+ x_2+ x_3= 50,$ (5.1)

$2x_1+3x_2+ x_3=270,$ (5.2)

$x_1+2x_2+3x_3= 80,$ (5.3)

TABLE I.

Place.	K_1' .	K_1'' .	K_1''' .	X_1 .	X_2 .	X_3 .
Unit.....	2.5	9.0	2.7	0	10	-4
First	-5.0	3.3	0.0	-0.7	10.9	-4.4
Second	-5.0	3.3	3.3	-0.79	11.00	-4.41
Third	-5.0	-3.3	6.7	-0.800	11.001	-4.401
Fourth	0.0	-6.7	3.3	-0.8000	11.0000	-4.4000
Fifth	0.0	0.0	0.0			
Final values : $x_1=-8$; $x_2=110$; $x_3=-44$.						

which, on dividing through by the coefficients of the leading terms, become

$x_1+0.5x_2+0.5x_3=25,$ (6.1)

$0.6x_1+ x_2+0.3x_3=90,$ (6.2)

$0.3x_1+0.6x_2+ x_3=26.6.$ (6.3)

The first iterative framework is

$X_1=2.5-0.5X_2-0.5X_3,$ (7.1)

$X_2=9.0-0.7X_1-0.3X_3,$ (7.2)

$X_3=2.7-0.3X_1-0.7X_2.$ (7.3)

(It will be noted that $n=1$.)

The value of Y , which is independent of the value of n , is now $0.25 \pm j0.227$, the modulus being 0.3, and three iterations suffice for each place. The manner in which the final results are approached is shown in Table I, the remainders disappearing at the fourth place.

It is not of course suggested that this method should be used in practice on equations such as (4), in which the constants are single figure numerics, but the procedure in more complicated cases will be evident from the illustration. As a general rule the convergency criterion is not used, as it is troublesome to work out with many unknowns. Instead, reliance is placed on the arrangement of the equations according to the rules given.

The usual practice in contracted arithmetic of taking 1.5 as 2, etc., where the result is to be given "to the nearest unit," may be followed in the working; the writer prefers, however, to take the lower value, which leads on the whole to slightly less arithmetic. The matter is not of vital moment, but whichever choice is adopted should be adhered to throughout.

An Extension of the Iterative Technique.

Reverting to the general principle of convergent iteration, it is seen to consist of the routine replacement in a pattern of one false value at a time by an improved value dependent on a systematic progression through the pattern, repeating the progression in the same sequence and following the same rule convention adopted at the beginning until the true pattern is resolved. It is, however, not binding to make this rule convention the replacement of $x_m^{(p-1)}$ by $x_m^{(p)}$ when iterating for $x_n^{(p)}$. Suppose we put $x_m^{(p')} = (1-A)x_m^{(p-1)} + Ax_m^{(p)}$ at and after any stage in the iterative process, using $x_m^{(p')}$ instead of $x_m^{(p)}$ in subsequent iterations. It is convenient to make A an arbitrary numerical constant. (It will be apparent that the orthodox process is a particular case in which A is unity). By proper choice of A it is often possible to obtain an iterative framework which converges more rapidly than the original.

Two main types of framework exist: (a) those in which the approach to the final values is always in the same sense, and (b) those in which the approach oscillates over the final value with diminishing amplitude. In addition, there are hybrids in which both characteristics appear, depending on the roots of the y equation. It is helpful to take A in case (a) as about 1.1 to 1.2, and in case (b) as 0.5 to 0.7. In the hybrid case it depends on the y roots which prevail in the later stages whether the convergence ultimately tends to (a) or (b), and in case of doubt it is better to take A as unity unless or until the position has been resolved.

The value of this device is naturally greater when a series of iterations to the same pattern is to be carried out, as in the method described in the first part of the paper, than when a single series, working throughout to the final number of places, is used. The way the first series works out serves as a guide to the choice of A for use in subsequent series. It is possible to shorten the process still further by changing A during the iteration, but it is doubtful whether such a course is really wise, in view of the extra complication.

The method outlined above, taking A as 0.5, has been applied to the

TABLE II.

<i>a.</i>	<i>b.</i>	<i>c.</i>	<i>d.</i>	<i>e.</i>	<i>f.</i>	<i>g.</i>	<i>h.</i>	<i>K.</i>	
1	-0.2356	-0.5625	0.5962	0.3646	-0.0925	-0.2059	0.2067	2.7083	(8.1)
0.4903	1	0.0044	-0.8224	0.2754	0.4712	0.0298	0.6208	2.9134	(8.2)
-0.0741	-0.3778	1	-1.0037	-0.0694	-0.1480	0.3805	-0.3644	-1.3975	(8.3)
0.1660	0.2813	0.0554	1	0.0744	0.1280	0.2378	0.4133	1.2667	(8.4)
-0.5368	-0.0146	0.1614	0.3357	1	-0.0305	-0.8718	-0.4544	0	(8.5)
-0.0163	-0.3811	0.1833	0.1379	0.0611	1	-0.6292	-0.2037	0	(8.6)
-0.0687	0.0689	-0.3024	-0.0245	0.1107	-0.1490	1	0.0339	0	(8.7)
0.0163	-0.0547	-0.1412	0.2921	0.0367	0.1721	0.4758	1	0	(8.8)

solution of the family of eight equations quoted by Schmidt* from Winny†. The constants, as arranged by the former for iteration to four places, are given in Table II. (The figure for g_2 in (8.2) is corrected for a printer's error in Schmidt's paper.) About six iterations are sufficient for each series, and give the results shown in Table III. The final line of results checks with the original K constants to within ± 0.0001 in each case, the average departure being about half this. The total labour involved is quite moderate.

TABLE III.

Place.	K_1' .	K_1'' .	K_1''' .	K_1'''' .	K_1''''' .	K_1'''''' .	K_1''''''' .	K_1'''''''' .
Unit...	2.7	2.9	-1.4	1.3	0	0	0	0
First ..	11.8	-0.7	-4.9	3.7	11.0	7.9	0.0	1.4
Second.	2.6	5.2	-1.8	2.0	4.4	11.3	-6.8	-3.9
Third ..	10.1	12.8	4.6	-1.3	8.9	-1.7	5.7	2.6
Fourth.	7.9	12.4	-1.9	8.8	9.7	-0.1	-8.7	-0.3
Fifth	All between 10 and -10.							
	X_1 .	X_2 .	X_3 .	X_4 .	X_5 .	X_6 .	X_7 .	X_8 .
Unit...	2	2	0	0	0	0	0	0
First ..	2.3	1.4	-0.2	0.3	1.2	0.4	0.0	0.1
Second.	2.33	1.42	-0.17	0.32	1.19	0.48	-0.05	0.08
Third ..	2.337	1.420	-0.171	0.315	1.207	0.481	-0.045	0.078
Fourth.	2.3378	1.4207	-0.1700	0.3153	1.2077	0.4807	-0.0457	0.0786
$x_m = X_m$ since $n=0$.								

Further Possibilities.

It is not essential in applying these methods to start *ab initio*. Approximate solutions obtained by any method may readily be further extended by the use of the process, providing only that the original data are sufficiently precise to make the extension worth while. In other words, the limitations are physical rather than mathematical.

It is of course possible to make the iterative framework still more approximate, *e. g.* by expressing the coefficients to the units place only, but this results in larger remainders, and it is believed that the degree of approximation suggested is the best compromise between over-elaborate figuring on the one hand and over-large remainders on the other.

* *Loc. cit.*

† Reports and Memo. of Royal Aero. Research Committee No. 1756.

In conclusion, it is suggested that the important self-checking property of a convergent iterative process is well worth retention. The methods described have been used by the writer in resolving a number of sets of equations relating to physical problems and involving four or five complex unknowns and five or six figure numerics; in no case has failure been experienced. The processes are, in fact, much easier to use than to describe, and it is hoped that other investigators may be tempted to give them a trial. The prime essential is to secure a reasonably convergent framework, and it is worth a little trouble at the start to secure this.

For a more lucid account of numerous details, and an admirable statement of general theory, the reader is referred to Schmidt's paper, to which the present writer is glad to acknowledge his indebtedness.

XLVII. *A Note on Stress Systems in Anisotropic Materials.*—I.

By A. E. GREEN*.

[Received October 9, 1942.]

1. Various problems of generalized plane stress distributions in anisotropic materials have been considered in a recent series of papers (see references). The materials were supposed to have two directions of symmetry and the stresses were derived from a stress function χ which satisfies the equation

$$s_{22} \frac{\partial^4 \chi}{\partial x^4} + (2s_{12} + s_{66}) \frac{\partial^4 \chi}{\partial x^2 \partial y^2} + s_{11} \frac{\partial^4 \chi}{\partial y^4} = 0. \quad (1.1)$$

It was assumed that this equation could be put in the form

$$\left(\frac{\partial^2}{\partial x^2} + \alpha_1 \frac{\partial^2}{\partial y^2} \right) \left(\frac{\partial^2}{\partial x^2} + \alpha_2 \frac{\partial^2}{\partial y^2} \right) \chi = 0, \quad (1.2)$$

where α_1 and α_2 were confined to real and positive values. The restriction that is implied by this assumption means that the roots of the quartic equation

$$s_{22} \rho^4 + (2s_{12} + s_{66}) \rho^2 + s_{11} = 0, \quad (1.3)$$

for ρ are all complex, and this condition is necessary in order that physically possible results may be obtained. It may be shown, however, that the equation (1.3) must, in fact, have complex roots because of the inequalities satisfied by the elastic constants as a result of the condition that the strain energy function

$$\frac{1}{2} s_{11} \widehat{xx}^2 + \frac{1}{2} s_{22} \widehat{yy}^2 + \frac{1}{2} s_{66} \widehat{xy}^2 + s_{12} \widehat{xx} \widehat{yy} \quad (1.4)$$

is always positive. Thus, equation (1.1) can always be expressed in the form (1.2), where α_1 and α_2 are real and positive.

* Communicated by the Author.

Plane stress systems in any homogeneous *æolotropic* material which satisfies Hooke's law, without any restrictions of symmetry of structure, can be derived from a stress function χ , which satisfies the more general equation

$$s_{22} \frac{\partial^4 \chi}{\partial x^4} - 2s_{26} \frac{\partial^4 \chi}{\partial x^3 \partial y} + (2s_{12} + s_{66}) \frac{\partial^4 \chi}{\partial x^2 \partial y^2} - 2s_{16} \frac{\partial^4 \chi}{\partial x \partial y^3} + s_{11} \frac{\partial^4 \chi}{\partial y^4} = 0. \quad (1.5)$$

The condition that this gives physically possible solutions for the types of problems considered in previous papers is that the equation

$$s_{22}\rho^4 + 2s_{26}\rho^3 + (2s_{12} + s_{66})\rho^2 + 2s_{16}\rho + s_{11} = 0 \quad (1.6)$$

for ρ has all complex roots, and in this general case it can be proved* that the roots are in fact complex because of the condition that the strain energy function is always positive.

2. Since the roots of (1.6) are complex the equation (1.5) can be expressed in the form

$$\left(\frac{\partial^2}{\partial x^2} + k_1 \frac{\partial^2}{\partial x \partial y} + \alpha_1 \frac{\partial^2}{\partial y^2} \right) \left(\frac{\partial^2}{\partial x^2} + k_2 \frac{\partial^2}{\partial x \partial y} + \alpha_2 \frac{\partial^2}{\partial y^2} \right) \chi = 0, \quad (2.1)$$

where $\alpha_1, \alpha_2, k_1, k_2$ are real, and where $\alpha_1 - \frac{1}{4}k_1^2, \alpha_2 - \frac{1}{4}k_2^2$ are positive. Also

$$\left. \begin{aligned} \alpha_1 + \alpha_2 + k_1 k_2 &= (2s_{12} + s_{66})/s_{22}, & \alpha_1 \alpha_2 &= s_{11}/s_{22}, \\ k_1 + k_2 &= -2s_{26}/s_{22}, & k_1 \alpha_2 + k_2 \alpha_1 &= -2s_{16}/s_{22}. \end{aligned} \right\} \quad (2.2)$$

For convenience real constants $\gamma_1, \gamma_2, \delta_1, \delta_2$ are defined by the equations

$$\left. \begin{aligned} \gamma_1 &= \frac{\alpha_1 - 1}{\alpha_1 + 1 + 2(\alpha_1 - \frac{1}{4}k_1^2)^{\frac{1}{2}}}, & \gamma_2 &= \frac{\alpha_2 - 1}{\alpha_2 + 1 + 2(\alpha_2 - \frac{1}{4}k_2^2)^{\frac{1}{2}}}, \\ \delta_1 &= \frac{-k_1}{\alpha_1 + 1 + 2(\alpha_1 - \frac{1}{4}k_1^2)^{\frac{1}{2}}}, & \delta_2 &= \frac{-k_2}{\alpha_2 + 1 + 2(\alpha_2 - \frac{1}{4}k_2^2)^{\frac{1}{2}}}, \end{aligned} \right\} \quad (2.3)$$

or by the equivalent forms

$$\left. \begin{aligned} \alpha_1 &= \frac{(1 + \gamma_1)^2 + \delta_1^2}{(1 - \gamma_1)^2 + \delta_1^2}, & \alpha_2 &= \frac{(1 + \gamma_2)^2 + \delta_2^2}{(1 - \gamma_2)^2 + \delta_2^2}, \\ \frac{k_1}{\alpha_1 - 1} &= -\frac{\delta_1}{\gamma_1}, & \frac{k_2}{\alpha_2 - 1} &= -\frac{\delta_2}{\gamma_2}. \end{aligned} \right\} \quad (2.4)$$

In some problems it is necessary to know the values of the displacements u and v . These are given by

$$\left. \begin{aligned} u &= (s_{12} - s_{11}) \frac{\partial \chi}{\partial x} + s_{11} \frac{\partial \psi}{\partial y} - s_{16} \frac{\partial \chi}{\partial y}, \\ v &= (s_{12} - s_{22}) \frac{\partial \chi}{\partial y} + s_{22} \frac{\partial \psi}{\partial x} - s_{26} \frac{\partial \chi}{\partial x}, \end{aligned} \right\} \quad (2.5)$$

where ψ has to be chosen so as to satisfy the equations

$$\frac{\partial^2 \psi}{\partial x \partial y} = \nabla_1^2 \chi, \quad (2.6)$$

* I am indebted to Prof. J. L. Burchall for a neat geometrical proof of this property.

$$\frac{\partial^2 \psi}{\partial x^2} + \alpha_1 \alpha_2 \frac{\partial^2 \psi}{\partial y^2} = \{(1 - \alpha_1)(1 - \alpha_2) - k_1 k_2\} \frac{\partial^2 \chi}{\partial x \partial y} - (k_1 + k_2) \frac{\partial^2 \chi}{\partial x^2} - (k_1 \alpha_2 + k_2 \alpha_1) \frac{\partial^2 \chi}{\partial y^2} \quad (2.7)$$

3. The most general solution of equation (2.1) may be taken ⁽³⁾ to be the real part of

$$\chi = f\{z + (\gamma_1 + i\delta_1)\bar{z}\} + g\{z + (\gamma_2 + i\delta_2)\bar{z}\}, \quad (3.1)$$

and the corresponding value of ψ , which satisfies (2.6) and (2.7), is such that

$$\left. \begin{aligned} \frac{\partial \psi}{\partial x} &= -\frac{4i(\gamma_1 + i\delta_1)}{1 - \gamma_1 - i\delta_1} f'\{z + (\gamma_1 + i\delta_1)\bar{z}\} - \frac{4i(\gamma_2 + i\delta_2)}{1 - \gamma_2 - i\delta_2} g'\{z + (\gamma_2 + i\delta_2)\bar{z}\}, \\ \frac{\partial \psi}{\partial y} &= \frac{4(\gamma_1 + i\delta_1)}{1 + \gamma_1 + i\delta_1} f'\{z + (\gamma_1 + i\delta_1)\bar{z}\} + \frac{4(\gamma_2 + i\delta_2)}{1 + \gamma_2 + i\delta_2} g'\{z + (\gamma_2 + i\delta_2)\bar{z}\}. \end{aligned} \right\} \quad (3.2)$$

The displacements (2.5) then take the form

$$\left. \begin{aligned} u &= (a_1 + ib_1) f'\{z + (\gamma_1 + i\delta_1)\bar{z}\} + (c_1 + id_1) g'\{z + (\gamma_2 + i\delta_2)\bar{z}\}, \\ v &= (a_2 + ib_2) f'\{z + (\gamma_1 + i\delta_1)\bar{z}\} + (c_2 + id_2) g'\{z + (\gamma_2 + i\delta_2)\bar{z}\}. \end{aligned} \right\} \quad (3.3)$$

where

$$\left. \begin{aligned} a_1 &= (1 + \gamma_1)(s_{12} - \alpha_2 s_{22}) + 2\delta_1 s_{22} L, & b_1 &= \delta_1(s_{12} - \alpha_2 s_{22}) + 2(1 - \gamma_1) s_{22} L, \\ c_1 &= (1 + \gamma_2)(s_{12} - \alpha_1 s_{22}) - 2\delta_2 s_{22} L, & d_1 &= \delta_2(s_{12} - \alpha_1 s_{22}) - 2(1 - \gamma_2) s_{22} L, \\ a_2 &= \delta_1(s_{12} - \alpha_1 s_{22}) + 2(1 + \gamma_1) s_{22} K, & b_2 &= (1 - \gamma_1)(s_{12} - \alpha_1 s_{22}) + 2\delta_1 s_{22} K, \\ c_2 &= \delta_2(s_{12} - \alpha_2 s_{22}) - 2(1 + \gamma_2) s_{22} K, & d_2 &= (1 - \gamma_2)(s_{12} - \alpha_2 s_{22}) - 2\delta_2 s_{22} K, \end{aligned} \right\} \quad (3.4)$$

and

$$\left. \begin{aligned} K &= \frac{\delta_1}{(1 - \gamma_1)^2 + \delta_1^2} - \frac{\delta_2}{(1 - \gamma_2)^2 + \delta_2^2}, \\ L &= \frac{\alpha_2 \delta_1}{(1 - \gamma_1)^2 + \delta_1^2} - \frac{\alpha_1 \delta_2}{(1 - \gamma_2)^2 + \delta_2^2}. \end{aligned} \right\} \quad (3.5)$$

4. The stress function corresponding to an isolated force at the origin, with components (X, Y) along the axes, is the real part of

$$\chi = (A + iB)\{z + (\gamma_1 + i\delta_1)\bar{z}\} \log \{z + (\gamma_1 + i\delta_1)\bar{z}\} + (C + iD)\{z + (\gamma_2 + i\delta_2)\bar{z}\} \log \{z + (\gamma_2 + i\delta_2)\bar{z}\}, \quad (4.1)$$

where A, B, C, D are real constants. In order that the corresponding displacements may be single-valued, these constants must satisfy the equations

$$\left. \begin{aligned} b_1 A + a_1 B + d_1 C + c_1 D &= 0, \\ b_2 A + a_2 B + d_2 C + c_2 D &= 0. \end{aligned} \right\} \quad (4.2)$$

Also, the components of force (X, Y) are found to be

$$\left. \begin{aligned} X &= 2\pi\{(1 - \gamma_1)A + (1 - \gamma_2)C + \delta_1 B + \delta_2 D\}, \\ Y &= -2\pi\{(1 + \gamma_1)B + (1 + \gamma_2)D + \delta_1 A + \delta_2 C\}. \end{aligned} \right\} \quad (4.3)$$

If an isolated force acts at the origin in any direction the stresses may be derived from the stress function (4.1) and the values of the constants A, B, C, D are found from equations (4.2) and (4.3).

5. When an isolated force acts at a point in the straight boundary of a semi-infinite plate it is not necessary to insist that the displacements are to be single-valued in describing any complete circuit round the point. The stress function is still given by (4.1) but the constants are now to be found from the condition that there is to be no stress along the straight boundary. Using polar co-ordinates (r, θ) , whose origin is at the point on the boundary, it may be seen that the stresses $r\theta$ and $\theta\theta$ are zero at every point in the plate, and in particular along the straight edge, if A, B, C, D satisfy the equations

$$\left. \begin{aligned} (1+\gamma_1)A + (1+\gamma_2)C - \delta_1 B - \delta_2 D &= 0, \\ -\delta_1 A - \delta_2 C + (1-\gamma_1)B + (1-\gamma_2)D &= 0. \end{aligned} \right\} \quad (5.1)$$

The components of force at the origin reduce in this case to

$$X = 2\pi(A+C), \quad Y = -2\pi(B+D). \quad (5.2)$$

Thus the problem of an isolated force at a point on the straight edge of a semi-infinite plate is solved by using the stress function (4.1), where the constants are now to be found from equations (5.1) and (5.2). The stress components $r\theta$ and $\theta\theta$ are zero at every point in the plate, and

$$\begin{aligned} \widehat{rr} = \frac{4}{r} \left\{ \frac{\{A(\gamma_1^2 + \delta_1^2 + \gamma_1) - B\delta_1\} \cos \theta + \{A\delta_1 - B(\gamma_1^2 + \delta_1^2 - \gamma_1)\} \sin \theta}{1 + \gamma_1^2 + \delta_1^2 + 2\gamma_1 \cos 2\theta + 2\delta_1 \sin 2\theta} \right. \\ \left. + \frac{\{C(\gamma_2^2 + \delta_2^2 + \gamma_2) - D\delta_2\} \cos \theta + \{C\delta_2 - D(\gamma_2^2 + \delta_2^2 - \gamma_2)\} \sin \theta}{1 + \gamma_2^2 + \delta_2^2 + 2\gamma_2 \cos 2\theta + 2\delta_2 \sin 2\theta} \right\}. \quad (5.3) \end{aligned}$$

6. Problems of stress distributions in an infinite plate which contains a circular hole may be solved by a method which is similar to that used in a previous paper⁽³⁾. The general stress function is taken to be

$$\begin{aligned} \chi = f\{z + (\gamma_1 + i\delta_1)\bar{z}\} + (A + iB)\{z + (\gamma_1 + i\delta_1)\bar{z}\} \log \{z + (\gamma_1 + i\delta_1)\bar{z}\} \\ + g\{z + (\gamma_2 + i\delta_2)\bar{z}\} + (C + iD)\{z + (\gamma_2 + i\delta_2)\bar{z}\} \log \{z + (\gamma_2 + i\delta_2)\bar{z}\}, \quad (6.1) \end{aligned}$$

where $f'\{z + (\gamma_1 + i\delta_1)\bar{z}\}$ and $g'\{z + (\gamma_2 + i\delta_2)\bar{z}\}$ have no poles at any points in the plate and where A, B, C, D are real constants which satisfy the equations (4.2) and (4.3). The remaining steps in the solution are exactly similar to those taken in the previous work, so that the details will not be given here. It is found that the stresses at the edge of the hole are still given by simple expressions.

If a uniform tension is applied to the plate at infinity parallel to the x -axis, then the stress at the edge of the circular hole is given by the equation

$$\begin{aligned} \theta\theta(1 + \gamma_1^2 + \delta_1^2 - 2\gamma_1 \cos 2\theta - 2\delta_1 \sin 2\theta)(1 + \gamma_2^2 + \delta_2^2 - 2\gamma_2 \cos 2\theta - 2\delta_2 \sin 2\theta) \\ = T\{(1 + \gamma_1)(1 + \gamma_2)(1 + \gamma_1 + \gamma_2 - \gamma_1\gamma_2 - 2 \cos 2\theta) + (\delta_1 + \delta_2)^2 - \delta_1^2\delta_2^2 \\ - \delta_1^2\gamma_2^2 - \delta_2^2\gamma_1^2 + 2\delta_1\delta_2 \cos 2\theta - 2(\delta_1 + \delta_2 + \delta_1\gamma_2 + \delta_2\gamma_1) \sin 2\theta\}. \quad (6.2) \end{aligned}$$

When the plate is acted on by a uniform shear S at infinity parallel to the axes of symmetry, the stress at the edge of the hole becomes

$$\begin{aligned} & \theta\theta(1+\gamma_1^2+\delta_1^2-2\gamma_1\cos 2\theta-2\delta_1\sin 2\theta)(1+\gamma_2^2+\delta_2^2-2\gamma_2\cos 2\theta-2\delta_2\sin 2\theta) \\ & = 4S\{\delta_1+\delta_2+(\gamma_1\gamma_2-\delta_1\delta_2-1)\sin 2\theta-(\gamma_1\delta_2+\gamma_2\delta_1)\cos 2\theta\}. \quad (6.3) \end{aligned}$$

If a uniform normal pressure P is applied to the edge of the hole then the circumferential stress takes the form

$$\begin{aligned} & \theta\theta(1+\gamma_1^2+\delta_1^2-2\gamma_1\cos 2\theta-2\delta_1\sin 2\theta)(1+\gamma_2^2+\delta_2^2-2\gamma_2\cos 2\theta-2\delta_2\sin 2\theta) \\ & = P\{1+(\gamma_1+\gamma_2)^2+(\delta_1+\delta_2)^2-3(\gamma_1^2+\delta_1^2)(\gamma_2^2+\delta_2^2) \\ & \quad + 2\{(\gamma_1+\gamma_2)(\gamma_1\gamma_2-\delta_1\delta_2-1)+(\delta_1+\delta_2)(\gamma_1\delta_2+\gamma_2\delta_1)\}\cos 2\theta \\ & \quad + 2\{(\gamma_1+\gamma_2)(\delta_1\gamma_2+\delta_2\gamma_1)-(\delta_1+\delta_2)(\gamma_1\gamma_2-\delta_1\delta_2+1)\}\sin 2\theta \\ & \quad - 2(\gamma_1\gamma_2-\delta_1\delta_2)\cos 4\theta-2(\gamma_1\delta_2+\gamma_2\delta_1)\sin 4\theta\}. \quad (6.4) \end{aligned}$$

Stress distributions in plates containing other types of holes have also been considered recently ⁽⁴⁾, and the work can be extended to apply to more general types of æolotropic materials by a similar process to that which is outlined above.

References.

- (1) Green, A. E., and Taylor, G. I., Proc. Roy. Soc. A, clxxiii. p. 162 (1939).
- (2) Green, A. E., Proc. Roy. Soc. A, clxxiii. p. 173 (1939).
- (3) Green, A. E., Proc. Roy. Soc. A, clxxx. p. 173 (1942).
- (4) Green, A. E. Not yet published.

XLVIII. A Note on Stress Systems in Æolotropic Materials.—II.

By A. E. GREEN *.

[Received October 9, 1942.]

1. PROFESSOR SEN † has given a simple method for solving problems of stress systems in a semi-infinite æolotropic plate which is bounded by a straight edge. His results, however, can only be applied to a very restricted class of æolotropic materials since a certain relation is supposed to be satisfied by the elastic constants, and this excludes, for example, such important highly æolotropic materials as spruce. In this note it is shown that Prof. Sen's method of solution may be extended so as to apply to any homogeneous æolotropic material which satisfies Hooke's law.

2. With the notation used in note I., it will be seen that a general stress function may be taken to be the real part of

$$\chi = f(z_1) + g(z_2), \quad \dots \dots \dots (2.1)$$

* Communicated by the Author.

† Phil. Mag. (7) xxvii. p. 596 (1939).

where
$$z_1 = \frac{z + (\gamma_1 + i\delta_1)\bar{z}}{1 + \gamma_1 + i\delta_1}, \quad z_2 = \frac{z + (\gamma_2 + i\delta_2)\bar{z}}{1 + \gamma_2 + i\delta_2} \quad (2.2)$$

The corresponding stresses are then the real parts of

$$\left. \begin{aligned} \widehat{xx} &= -\frac{(1 - \gamma_1 - i\delta_1)^2}{(1 + \gamma_1 + i\delta_1)^2} f''(z_1) - \frac{(1 - \gamma_2 - i\delta_2)^2}{(1 + \gamma_2 + i\delta_2)^2} g''(z_2), \\ \widehat{yy} &= f''(z_1) + g''(z_2), \\ \widehat{xy} &= -\frac{i(1 - \gamma_1 - i\delta_1)}{1 + \gamma_1 + i\delta_1} f''(z_1) - \frac{i(1 - \gamma_2 - i\delta_2)}{1 + \gamma_2 + i\delta_2} g''(z_2). \end{aligned} \right\} \quad (2.3)$$

If the boundary is to be subjected to normal stress only, then $\widehat{xy} = 0$ when $y = 0$, and the stresses (2.3) may be taken to be the real parts of

$$\left. \begin{aligned} \widehat{xx} &= -\left(\frac{1 - \gamma_1 - i\delta_1}{1 + \gamma_1 + i\delta_1} \right) \left\{ \frac{1 - \gamma_1 - i\delta_1}{1 + \gamma_1 + i\delta_1} f''(z_1) - \frac{1 - \gamma_2 - i\delta_2}{1 + \gamma_2 + i\delta_2} f''(z_2) \right\}, \\ \widehat{yy} &= f''(z_1) - \frac{(1 - \gamma_1 - i\delta_1)(1 + \gamma_2 + i\delta_2)}{(1 + \gamma_1 + i\delta_1)(1 - \gamma_2 - i\delta_2)} f''(z_2), \\ \widehat{xy} &= -\frac{i(1 - \gamma_1 - i\delta_1)}{1 + \gamma_1 + i\delta_1} \{ f''(z_1) - f''(z_2) \}. \end{aligned} \right\} \quad (2.4)$$

Also, on the boundary

$$(\widehat{yy})_{y=0} = \frac{2(\gamma_1 - \gamma_2 + i\delta_1 - i\delta_2)}{(1 + \gamma_1 + i\delta_1)(1 - \gamma_2 - i\delta_2)} f''(z), \quad (2.5)$$

where only the real part of the right-hand side is to be used.

If there is to be only a distribution of shear stress along the boundary $y = 0$, then the stresses are the real parts of

$$\left. \begin{aligned} \widehat{xx} &= -\frac{(1 - \gamma_1 - i\delta_1)^2}{(1 + \gamma_1 + i\delta_1)^2} f''(z_1) + \frac{(1 - \gamma_2 - i\delta_2)^2}{(1 + \gamma_2 + i\delta_2)^2} f''(z_2), \\ \widehat{yy} &= f''(z_1) - f''(z_2), \\ \widehat{xy} &= -\frac{i(1 - \gamma_1 - i\delta_1)}{1 + \gamma_1 + i\delta_1} f''(z_1) + \frac{i(1 - \gamma_2 - i\delta_2)}{1 + \gamma_2 + i\delta_2} f''(z_2), \end{aligned} \right\} \quad (2.6)$$

and the shear stress on the boundary is the real part of

$$(\widehat{xy})_{y=0} = \frac{2i(\gamma_1 - \gamma_2 + i\delta_1 - i\delta_2)}{(1 + \gamma_1 + i\delta_1)(1 + \gamma_2 + i\delta_2)} f''(z). \quad (2.7)$$

3. As an example, one of the problems which was solved by Prof. Sen will be reconsidered here.

Suppose that on the boundary

$$\left. \begin{aligned} (\widehat{xy})_{y=0} &= 0, \\ (\widehat{yy})_{y=0} &= -p \text{ when } -a < x < a, \\ &= 0 \text{ for other values of } x. \end{aligned} \right\} \quad (3.1)$$

It will then be seen from (2.5) that

$$f''(z) = \frac{p(1 + \gamma_1 + i\delta_1)(1 - \gamma_2 - i\delta_2)}{2\pi(\gamma_1 - \gamma_2 + i\delta_1 - i\delta_2)} i \log \frac{z - a}{z + a}. \quad (3.2)$$

Hence, using (2.4), the stresses may be taken to be the real parts of

$$\left. \begin{aligned} \widehat{xx} &= \frac{ip(1-\gamma_1-i\delta_1)(1-\gamma_2-i\delta_2)}{2\pi(\gamma_1-\gamma_2+i\delta_1-i\delta_2)} \left\{ \frac{1-\gamma_2-i\delta_2}{1+\gamma_2+i\delta_2} \log \frac{z_2-a}{z_2+a} \right. \\ &\quad \left. - \frac{1-\gamma_1-i\delta_1}{1+\gamma_1+i\delta_1} \log \frac{z_1-a}{z_1+a} \right\}, \\ \widehat{yy} &= \frac{ip}{2\pi(\gamma_1-\gamma_2+i\delta_1-i\delta_2)} \left\{ (1+\gamma_1+i\delta_1)(1-\gamma_2-i\delta_2) \log \frac{z_1-a}{z_1+a} \right. \\ &\quad \left. - (1-\gamma_1-i\delta_1)(1+\gamma_2+i\delta_2) \log \frac{z_2-a}{z_2+a} \right\}, \\ \widehat{xy} &= \frac{p(1-\gamma_1-i\delta_1)(1-\gamma_2-i\delta_2)}{2\pi(\gamma_1-\gamma_2+i\delta_1-i\delta_2)} \left\{ \log \frac{z_1-a}{z_1+a} - \log \frac{z_2-a}{z_2+a} \right\}. \end{aligned} \right\} \quad (3.3)$$

When a concentrated pressure P is applied at a point on the boundary the resulting stresses may be regarded as the limiting values of expressions (3.3) when $a \rightarrow 0$ and $ap \rightarrow P$, or they may be found independently from (2.4) and (2.5) by the same process as that used by Prof. Sen, or they may be found by the slightly different method which was given in Note I. (p. 416).

With the help of the present analysis, the results of the other problems which were considered by Prof. Sen can also be extended so as to apply to a general homogeneous æolotropic material. In addition, the problem of an isolated force which acts at an internal point of the plate may also be solved by the present method. This problem was considered in a previous paper*, by another method, for æolotropic materials which have two directions of symmetry at right angles.

XLIX. *Proofs of some Formulæ for the Hypergeometric Function and the E-function.*

By T. M. MACROBERT, Professor of Mathematics, University of Glasgow †.

[Received November 23, 1942.]

§ 1. *Introductory.*

In section 2 of this paper connexions between various formulæ for the hypergeometric function are discussed. Section 3 contains a proof of an E-function formula, and in section 4 a known integral involving modified Bessel Functions is deduced from this formula.

§ 2. *Proofs of some Formulæ for the Hypergeometric Function.*

The formula (Gauss, *Ges. Werke*, iii. p. 226)

$$F\left(\begin{matrix} 2\alpha, 2\beta; x \\ \alpha+\beta+\frac{1}{2} \end{matrix}\right) = F\left\{\begin{matrix} \alpha, \beta; 4x(1-x) \\ \alpha+\beta+\frac{1}{2} \end{matrix}\right\} \quad \dots \quad (1)$$

* Proc. Roy. Soc. A, clxxiii. p. 173 (1939).

† Communicated by the Author.

can be verified by expressing the term containing x^n on the R.H.S. in the form

$$\frac{(\alpha; n)(\beta; n)}{n! (\alpha + \beta + \frac{1}{2}; n)} (4x)^n {}_3F_2 \left(\begin{matrix} -\frac{1}{2}n, \frac{1}{2} - \frac{1}{2}n, \frac{1}{2} - \alpha - \beta - n \\ 1 - \alpha - n, 1 - \beta - n \end{matrix}; 1 \right).$$

where $(k; n) = k(k+1) \dots (k+n-1)$,

and then applying Saalschütz's Theorem

$${}_3F_2 \left(\begin{matrix} a, b, c \\ d, e \end{matrix}; 1 \right) = \frac{\Gamma(d)\Gamma(1+a-e)\Gamma(1+b-e)\Gamma(1+c-e)}{\Gamma(1-e)\Gamma(d-a)\Gamma(d-b)\Gamma(d-c)}, \quad (2)$$

where a, b or c is a negative integer and $d+e=a+b+c+1$.

If β is replaced by $\alpha + \frac{1}{2}$, formula (1) reduces to

$$(1-x)^{-2\alpha} = F \left\{ \begin{matrix} \alpha, \alpha + \frac{1}{2}; 4x(1-x) \\ 2\alpha + 1 \end{matrix} \right\}, \quad (3)$$

$$= (1-2x) F \left\{ \begin{matrix} \alpha + 1, \alpha + \frac{1}{2}; 4x(1-x) \\ 2\alpha + 1 \end{matrix} \right\}. \quad (4)$$

Now, in (4), replace x by $\frac{1}{2}\{1 - \sqrt{1-x}\}$ and α by $\frac{1}{2}(\alpha-1)$, and it becomes

$$\frac{1}{\sqrt{1-x}} \left\{ \frac{2}{1 + \sqrt{1-x}} \right\}^{\alpha-1} = F \left(\frac{1}{2}\alpha, \frac{1}{2}\alpha + \frac{1}{2}; \alpha; x \right), \quad (5)$$

(cf. Phil. Mag. ser. 7, xxvi. p. 86 (1938)).

Next, in (1), replace α, β and x by $\frac{1}{2}\alpha, \frac{1}{2} + \frac{1}{2}\alpha - \beta$ and $z/(z-1)$. Then

$$F \left\{ \begin{matrix} \frac{1}{2}\alpha, \frac{1}{2} + \frac{1}{2}\alpha - \beta; -\frac{4z}{(1-z)^2} \\ 1 + \alpha - \beta \end{matrix} \right\} = F \left\{ \begin{matrix} \alpha, 1 + \alpha - 2\beta; \frac{z}{z-1} \\ 1 + \alpha - \beta \end{matrix} \right\},$$

from which it follows that

$$F \left(\begin{matrix} \alpha, \beta; z \\ 1 + \alpha - \beta \end{matrix} \right) = (1-z)^{-\alpha} F \left\{ \begin{matrix} \frac{1}{2}\alpha, \frac{1}{2} + \frac{1}{2}\alpha - \beta; -\frac{4z}{(1-z)^2} \\ 1 + \alpha - \beta \end{matrix} \right\}, \quad (6)$$

a formula also due to Gauss (*Ges. Werke*, iii. p. 225. See also Phil. Mag. ser. 7, xxvi. p. 86).

Again, apply (1) to the R.H.S. of the identity

$$F \left(\begin{matrix} -n, n+1; \frac{1-x}{2} \\ m+1 \end{matrix} \right) = \left(\frac{1+x}{2} \right)^m F \left(\begin{matrix} m+n+1, m-n; \frac{1-x}{2} \\ m+1 \end{matrix} \right),$$

and so obtain the formula

$$F \left(\begin{matrix} -n, n+1; \frac{1-x}{2} \\ m+1 \end{matrix} \right) = \left(\frac{1+x}{2} \right)^m F \left(\begin{matrix} \frac{m+n+1}{2}, \frac{m-n}{2}; 1-x^2 \\ m+1 \end{matrix} \right), \quad (7)$$

from which it follows that

$$F \left(\begin{matrix} -n, n+1; \frac{1-x}{2} \\ m+1 \end{matrix} \right) = \left(\frac{1+x}{2x} \right)^m x^n F \left(\begin{matrix} \frac{m-n+1}{2}, \frac{m-n}{2}; 1-\frac{1}{x^2} \\ m+1 \end{matrix} \right). \quad (8)$$

The last formula is equivalent to Whipple's Transformation (*Proc. Lond. Math. Soc.* xvi. p. 301 (1916)).

§ 3. An E-function Formula.

In a recent paper (Quart. Journ. of Math. xiii. p. 67 (1942)) an integral involving a product of two E-functions was evaluated in the form

$$\int_0^\infty e^{-ty^{y-1}} E(\alpha, \beta : : yt) E(\lambda, \mu : : zt) dt \\ = \sum_{\alpha, \beta} \frac{\pi \Gamma(\lambda) \Gamma(\mu)}{\sin(\beta - \alpha) \pi} \sum_{r=0}^\infty \frac{\Gamma(\alpha+r) \Gamma(\alpha+\gamma+\lambda+r) \Gamma(\alpha+\gamma+\mu+r)}{r! \Gamma(\alpha-\beta+1+r) \Gamma(\alpha+\gamma+\lambda+\mu+r)} \\ \times y^{\alpha+r} F\left(\begin{matrix} \lambda, \mu; \\ \alpha+\gamma+\lambda+\mu+r \end{matrix}; \frac{z-1}{z}\right), \quad (9)$$

where $|y| < 1$, $R(z) > \frac{1}{2}$, $R(\alpha+\gamma+\lambda) > 0$, $R(\alpha+\gamma+\mu) > 0$, $R(\beta+\gamma+\lambda) > 0$, $R(\beta+\gamma+\mu) > 0$.

The R.H.S. of (9) may be written

$$\sum_{\alpha, \beta} \frac{\pi \Gamma(\lambda)}{\sin(\beta - \alpha) \pi} \sum_{r=0}^\infty \frac{\Gamma(\alpha+r) \Gamma(\alpha+\gamma+\mu+r)}{r! \Gamma(\alpha-\beta+1+r)} y^{\alpha+r} \\ \times \int_0^1 t^{\mu-1} (1-t)^{\alpha+\gamma+\lambda+r-1} \left(1-t \frac{z-1}{z}\right)^{-\lambda} dt \\ = \Gamma(\lambda) \int_0^1 t^{\mu-1} (1-t)^{\gamma+\lambda-1} \left(1-t \frac{z-1}{z}\right)^{-\lambda} \\ \times \sum_{\alpha, \beta} \{y(1-t)\}^\alpha \Gamma(\alpha) \Gamma(\beta-\alpha) \Gamma(\alpha+\gamma+\mu) F\left\{\begin{matrix} \alpha, \alpha+\gamma+\mu; \\ \alpha-\beta+1 \end{matrix}; y(1-t)\right\} dt.$$

Hence

$$\int_0^\infty e^{-ty^{y-1}} E(\alpha, \beta : : yt) E(\lambda, \mu : : zt) dt \\ = \Gamma(\alpha) \Gamma(\beta) \Gamma(\lambda) \Gamma(\alpha+\gamma+\mu) \Gamma(\beta+\gamma+\mu) \{ \Gamma(\alpha+\beta+\gamma+\mu) \}^{-1} y^\alpha \\ \times \int_0^1 t^{\mu-1} (1-t)^{\alpha+\gamma+\lambda-1} \left(1-t \frac{z-1}{z}\right)^{-\lambda} F\left\{\begin{matrix} \alpha, \alpha+\gamma+\mu; \\ \alpha+\beta+\gamma+\mu \end{matrix}; 1-y(1-t)\right\} dt, \quad (10)$$

where $0 < y < 1$, $R(z) > \frac{1}{2}$, $R(\alpha+\gamma+\lambda) > 0$, $R(\alpha+\gamma+\mu) > 0$, $R(\beta+\gamma+\lambda) > 0$, $R(\beta+\gamma+\mu) > 0$, $R(\mu) > 0$.

On making $y \rightarrow 1$, expanding the hypergeometric function and integrating, formula (10) of the paper referred to above is obtained.

§ 4. Derivation of a Modified Bessel Function Integral from an E-function Integral.

The formula (Titchmarsh, Journ. of Lond. Math. Soc. ii. p. 98 (1926))

$$\int_0^\infty t^{l-1} K_m(xt) K_n(t) dt \\ = \Gamma\left(\frac{l+m+n}{2}\right) \Gamma\left(\frac{l+m-n}{2}\right) \Gamma\left(\frac{l-m+n}{2}\right) \Gamma\left(\frac{l-m-n}{2}\right) 2^{l-3} \\ \times x^{-l-n} [\Gamma(l)]^{-1} F\left(\frac{l+m+n}{2}, \frac{l-m+n}{2}; l; 1-x^{-2}\right), \quad (11)$$

where $R(x) > -1$, $|1-x^{-2}| < 1$, $R(l \pm m \pm n) > 0$, can be deduced from formula (10) by means of the equation

$$\cos n\pi E\left(\frac{1}{2}+n, \frac{1}{2}-n : : 2x\right) = \sqrt{(2\pi x)e^x K_n(x)}. \quad (12)$$

In (10) put $\alpha = \frac{1}{2} + m$, $\beta = \frac{1}{2} - m$, $\lambda = \frac{1}{2} + n$, $\mu = \frac{1}{2} - n$, $\gamma = l - 1$, $y = 2x/(x+1)$, $z = 2/(x+1)$, and replace t on the left by $t(x+1)$. Then, on applying formula (12), it is found that

$$\begin{aligned} \int_0^\infty t^{l-1} K_m(xt) K_n(t) dt &= \frac{\pi \Gamma(l+m-n) \Gamma(l-m-n)}{\Gamma(\frac{1}{2}-n) \Gamma(l-n+\frac{1}{2})} \frac{2^{m-\frac{1}{2}} x^m}{(x+1)^{l+m-\frac{1}{2}}} \\ &\times \int_0^1 (1-t)^{l+m+n-1} \left(t - t^2 \frac{1-x}{2} \right)^{-\frac{1}{2}-n} \\ &\times F \left\{ \begin{matrix} \frac{1}{2}+m, l+m-n; \\ l-n+\frac{1}{2} \end{matrix} ; 1 - \frac{2x}{x+1} (1-t) \right\} dt, \end{aligned}$$

with suitable conditions for convergence.

Now, in the latter integral, put

$$1-y = \left(t - t^2 \frac{1-x}{2} \right) \frac{2}{1+x},$$

and it reduces to

$$\begin{aligned} \left(\frac{x+1}{2} \right)^{l+m-\frac{1}{2}} x^{-l-m-n} \int_0^1 \left\{ \frac{2y}{1+\sqrt{(1-\lambda)}} \right\}^{l+m+n-1} (1-y)^{-n-\frac{1}{2}} \frac{1}{\sqrt{(1-\lambda)}} \\ \times F \left\{ \begin{matrix} \frac{1}{2}+m, l+m-n; \\ l-n+\frac{1}{2} \end{matrix} ; 1 - \frac{2y}{1+\sqrt{(1-\lambda)}} \right\} dy, \end{aligned}$$

where $\lambda = (1-x^{-2})y$.

Again, in this integral, expand the hypergeometric function and apply formula (5). Then it becomes

$$\begin{aligned} \sum_{r=0}^\infty \frac{(\frac{1}{2}+m; r)(l+m-n; r)}{r! (l-n+\frac{1}{2}; r)} \sum_{s=0}^r (-1)^s {}^r C_s \\ \times \int_0^1 y^{l+m+n+s-1} (1-y)^{-n-\frac{1}{2}} \\ \times F \left\{ \begin{matrix} \frac{l+m+n+s}{2}, \frac{l+m+n+s+1}{2}; \\ l+m+n+s \end{matrix} ; (1-x^{-2})y \right\} dy. \end{aligned}$$

Now the integral in the last line is equal to

$$\frac{\Gamma(l+m+n+s) \Gamma(\frac{1}{2}-n)}{\Gamma(l+m+\frac{1}{2}+s)} F \left\{ \begin{matrix} \frac{l+m+n+s}{2}, \frac{l+m+n+s+1}{2}; \\ l+m+\frac{1}{2}+s \end{matrix} ; 1-x^{-2} \right\}.$$

Hence, on changing the order of summation, we find that

$$\int_0^\infty t^{l-1} K_m(xt) K_n(t) dt = \frac{\pi \Gamma(l+m-n) \Gamma(l-m-n)}{\Gamma(l-n+\frac{1}{2}) 2^l x^{l+n}} \times J,$$

where

$$J = \sum_{r=0}^{\infty} \frac{(\frac{1}{2}+m; r)(l+m-n; r)}{r! (l-n+\frac{1}{2}; r)} \sum_{p=0}^{\infty} \frac{\Gamma(l+m+n+2p)}{p! \Gamma(l+m+\frac{1}{2}+p)} \\ \times \left(\frac{1-x^{-2}}{4} \right)^p F \left(\begin{matrix} -r, l+m+n+2p; 1 \\ l+m+\frac{1}{2}+p \end{matrix} \right).$$

Here apply Gauss's Theorem, and again change the order of summation; thus

$$J = \sum_{p=0}^{\infty} \frac{\Gamma(l+m+n+2p)}{p! \Gamma(l+m+\frac{1}{2}+p)} \left(\frac{1-x^{-2}}{4} \right)^p F \left(\begin{matrix} l+m-n, \frac{1}{2}+m, \frac{1}{2}-n-p; 1 \\ l-n+\frac{1}{2}, l+m+\frac{1}{2}+p \end{matrix} \right).$$

On applying Dixon's Formula* and the duplication formula for the Gamma Function, formula (11) is obtained.

L. Additional Note on the Trichromatic Theory of Colour Vision.

(Referring to Phil. Mag., Aug. 1942, p. 559.)

By Prof. W. PEDDIE †.

[Received November 20, 1942.]

HELMHOLTZ's statement that in general there are three independent activities at work in visual perception is absolutely correct, and entirely unavoidable. It is simply a statement of the essence of *trichromacy*. It asserts that the great majority of human eyes are possessed of three colour freedoms; so that three data of specification and measurement are in general needed to remove these freedoms from external illumination and make definite the character of the visual activities which are being studied. It is now understood that the statement refers to cone stimulation.

The fact of trichromacy makes each of the three externally originated luminous stimuli (red, green, blue—whatever they be called) as absolutely independent as are a northward, an eastward, and an upward displacement in the specification of motion. No component of any one of the three standards is present in either of the other two. That is what Helmholtz's word "independent" means, and must mean in order to make progress possible.

$$* F \left(\begin{matrix} a, b, c; 1 \\ a-b+1, a-c+1 \end{matrix} \right)$$

$$= \frac{(1+\frac{1}{2}a)\Gamma(1+a-b)\Gamma(1+a-c)\Gamma(1+\frac{1}{2}a-b-c)}{\Gamma(1+a)\Gamma(1+\frac{1}{2}a-b)\Gamma(1+\frac{1}{2}a-c)\Gamma(1+a-b-c)},$$

where $1+\frac{1}{2}a > b+c$.

† Communicated by the Author,

Luminous Stimuli.

By "luminous" stimuli we normally mean "visible to the eye." But we have to consider cases in which, though actually present, a stimulus may be too feeble to give rise to perception directly. For example, the arbitrary blue standard may be entirely absent, and so also at first the green, while the red is just not powerful enough to cause the sensation of vision. A very slight, in itself imperceptible, addition of the green standard may now bring the red to perception. In this way there arises the immense value of the idea of the threshold; and with it also the idea of the physiological unit as the least perceptible stimulus—consecutive just perceptible steps being (arbitrarily) considered to be equal. Some day perhaps we shall have sufficiently delicate apparatus to measure, as energy, these small quantities with good accuracy, though knowledge of the type of the internally transmitted energy might still be unattainable, and therefore knowledge of the corresponding "Joule's Equivalent" also be wanting. However that may be, the bringing of invisible red light to perception by means of adding a still less visible amount of green light to it does not give an example of interference as destructive of Helmholtz's principle of independence. His "independent activities" are *externally* started. Other activities are of internal origin, and must be carefully separated from the three external activities which produce them and may affect them. It might be extremely misleading to use the term interference with regard to them as very likely to lead to the idea that Helmholtz's *treatment* of the subject is wrong.

Fusion of Effects.

Very careful consideration must be given to the question of the activities when they are internal. Young's original, tentatively suggested, isolated paths suit, in so far as they give a triple system. So also does Helmholtz's tentatively suggested triplicity of chemical substances. But Helmholtz recognized that *fusion* of effects was required to suit later found phenomena of dichromasy. Similarly, less complete fusions, but nevertheless fusions, are needed to explain the phenomena of anomalous trichromasy. These fusions are of internal origin.

Helmholtz himself, when dealing with the necessity for substitution of the fusion of effects instead of the paralysis of one centre or circuit, was essentially dealing with the question of cross connexions such as are exemplified in the electrical model shown in the previous paper. But, most certainly, he did not therewith give up the characteristic of independence of the several activities. He pointed out in detail how partial substitution of a green responsive substance by a red responsive one, or conversely, could give rise to the fusion of effects. This simple hypothetical process establishes a cross connexion. The electrical model establishes it by shunt conductors.

We could as easily use a model consisting of tubes in which an incompressible liquid flowed, the tubes also being cross connected. By making

alterations in the various channels, we could illustrate the use of a variety of trichromatic systems.

In particular, two activities may have their results compulsorily equalized because of a peculiarity in the mechanism which carries the activities. Any such changes are due to *guidance* of the activities by alterations in the channels which carry them. Damping of the magnitude of an effect may also appear. No doubt the term "interference" could be applied to either of these actions. But the words "shunting" or "damping," or some such terms, would, if used instead, avoid the unnecessary appearance of antagonism to Helmholtz's Principle.

Time Variations.

As important is the fact that we could illustrate the process of change from one system to another; and that not merely instantaneously, but as progressing *in time*. Illustrations are given in 'Colour Vision,' 1922, pp. 176-86. All problems of after images, of their origination, and of their decay, come within the scope of the Young-Helmholtz Theory.

Representative Models.

For further elucidation we may consider the flow of a medium consisting of small smooth coloured balls through three main channels with cross-connecting channels. Let balls of one colour alone enter each main channel; balls of all three standard colours, in different proportions determinable by the channel characteristics, would reach the exit ends. There is complete invariability in the number of balls of any one kind crossing any complete section of the complex channel. The dependence on the nature of the cross connexions illustrates well the utter impossibility of settling a question regarding the absolute fundamentals by experiments on colour mixture alone.

So long as the conditions at entry, and therefore the values of the external variables, do not change, the flow goes on *in complete independence* of any question of the entire separation of the three main channels, or any question of some part of the complete compound channel carrying a mixture of the three types of particle.

But if, keeping the rates of entry as before, that is, keeping the external variables fixed, we alter the conditions of the channels, the corresponding changes at the exit ends are initiated. With the accession of cross-connected channels, the model has ceased to be discriminative of the absolute fundamentals. To make it still more completely illustrative of visual actions, we might endow the use of an entry with an influence on the cross connexions, or even on the rates of entry themselves.

It is only a model in which the three paths are completely isolated which can illustrate the absolute fundamental actions. If it be the case that development of the visual organ has proceeded from total fusion towards total isolation of paths, we can understand a final attainment to a visual spectrum proceeding from red to green, through red-greens,

and from green to blue through green-blues, yellows having ceased to be perceived.

The Law of Coefficients, and the Physiological Unit.

The former name was given by v. Kries to a statement due to Fechner and Helmholtz. It may be put into various forms. Dr. W. D. Wright (Proc. Roy. Soc. vol. liv, p. 303, 1942) has used it in an extended form in connexion with binocular tests; and it has been employed also by Mr. H. V. Walters in a continuation of these tests by a different experimental process (Proc. Roy. Soc. B, vol. cxxxi, p. 27, 1942). Dr. Wright has pointed out (Proc. Roy. Soc. B, vol. cxv, p. 9, 1934) the analogy of this law to the Weber-Fechner law. For the present purpose it may be used in the restricted form :—The ratio of the intensities required to cause a given sensation, respectively without and with fatigue, is constant. Limiting the difference of intensity to that which is just perceptible, we find

$$dI/I = \text{const.}$$

if the law of coefficients be true. But this becomes

$$dI/I = kdS,$$

if Fechner's law be accepted in its simplest form. So, if both laws be true, all just perceptible steps in sensation are equal. Thus the approximate truth of both laws may be regarded as furnishing a *proof* of the approximate constancy of the physiological unit of sensation.

So, integrating, we conversely bring in, as the mathematical constant of integration, the fundamentally important, physical, physiological, and psychological threshold of perception—provided that all just perceptible steps in sensation are equal. It then belongs to the field of experiment to determine upon what variables the threshold is dependent, and the manner of its dependence.

There is no need for the employment of three laws—the psychological law, the coefficient law, and the equi-unitary law. Only two can be independent. The last is the simplest of all to express: the first is the most fundamental in requirement. And the threshold seems to be too important to be omitted from expression.

The Internal Variables.

It is not possible to urge too strongly that a symbolic representation should be used which is *quite independent of any particular postulated process* of physical or psychological action involved—unless the primary object of investigation be actually the nature of the processes. This is furnished by the psycho-physical law. Whatever be the true form of that law, all results must be in accordance with it. And Fechner's law, in its modified form as given by Helmholtz, seems, in consequence of modern work, to have very strong support. Further, it is still capable of expansion.

Now the internal variables come into play through influences which affect the conditions of the channels and may enter in as a consequence

of the magnitudes of the external stimuli x, y, z . Other external influences, *e. g.*, pressure on the eyeball, may originate them. They may increase or diminish with time, as during a process of alteration of accommodation. In short, they may depend on any condition which affects physiological or psychological states; or even anatomical conditions, as in brain injuries.

No one of these actions has the slightest power of altering the magnitude of an external variable, though care has to be taken with regard to pupil aperture. But they may, and generally do, alter the responses; and therefore change the thresholds. Thus the experimental results are capable of formal representation by expressing the thresholds as functions of the quantities upon which they are dependent. So, " a " being the threshold value of the stimulus x , we may write

$$x = a \cdot u(x, y, z, t, \dots, l, m, \dots, p, q, \text{etc.} \dots),$$

where u is the required function; t is the time; l, m, n , are the wavelengths of the three external fundamentals; while p, q , etc., take account of any other influences. In the determination of $u(\dots)$, the experimenter varies one quantity at a time. Similarly the expressions for y and z are found.

If the psychological law adopted be the unmodified one of the Young-Helmholtz theory, we have, as the expression for the Sensation,

$$S = k \cdot \log u(\dots)$$

to a first approximation.

Conclusion and Appeal.

The most modern work on colour vision, carried out now in Physical, in Physiological, and in Psychological Laboratories, seems to be leading more and more completely to recognition of the satisfactory position of the Trichromatic Theory of Colour Vision in regard to its main postulates. The correctness of the *trichromatic* aspect itself has become fully recognized. New directions of exploration are now becoming possible. Still more will open out as advancement grows. And that advance will be hastened—just as advances in all branches of science are aided and made more easy—by the use of a common plan for formulation and a common language for expression.

As at the conclusion of my previous paper, so now at the end of this additional note, I wish to express the hope that the extended use of the Young-Helmholtz Theory, as an aid to research, will meet with increasingly full reward.

My main aim is only that of securing, as far as may be possible, in our scientific work on the subject of colour sensation, the advantage of a treatment which, besides appearing to be the simplest possible, will bring about that saving of time and labour to many workers which arises from uniformity of records. Another is to give a short account of the history and present position of the trichromatic theory, from the physical point of view, to workers who approach the subject from other directions.

LI. *Notices respecting New Books.*

Bibliography of the Literature relating to Constitutional Diagrams of Alloys.

By Dr. J. L. HAUGHTON. [Pp. iv+163.] (London, 1942. The Institute of Metals, 4 Grosvenor Gardens, S.W. 1. Price 3s. 6d. post free.)

WITH increasing specialization—and in almost every branch of science, data are accumulating at a rate which always astonishing and sometimes terrifying—there comes a growing need for trustworthy guides through these mazes of facts, figures and theories which are provided by the contributors of papers to the proceedings of the scientific societies concerned.

One's own bibliography is usually the best for the particular user, but many of us do not possess the art of the bibliographer, and are correspondingly grateful for the unselfish labours of those who can lay claim to a skill that we do not possess, and workers in the metallographic and allied fields are deeply indebted to Dr. Haughton for the very comprehensive bibliography which he has just published. It is no small feat to have put in order more than 5000 references to "papers dealing with the constitution of binary, ternary, and higher alloy systems, both ferrous and non-ferrous. References have been included not only to papers of a purely constitutional character but also to many X-ray and physical-property studies of alloys, which have some bearing on the constitution. To furnish some guide amongst the references (which in a few systems exceed 100), asterisks have been placed against those in the originals of which a new equilibrium diagram, or portion of one, is to be found."

The bibliography includes references to all papers seen up to the time of going to press and should, indeed, prove invaluable to all interested in the study of the constitution of metallic alloys. It is clear and full, well produced, and may be obtained at a remarkably modest price. *O si sic omnes!*

Table of the Coefficients of Everett's Central Difference Interpolation Formula.

By A. J. THOMPSON. Second Edition. (*Tracts for Computers*, edited by E. S. PEARSON. No. V.) [Pp. viii+32.] (Cambridge University Press, 1943. Price 5s.)

IF differences are to be tabulated for use with interpolation formulæ, the Everett formula is advantageous as only even order differences are required. This interpolation formula was the first whose coefficients were published at interval 0.001, in the first edition of this tract (1921). That edition has been out of print for some time. In the new edition the main twenty page table is more conveniently arranged in that the coefficients up to those for the eighth differences appear on one line at a single opening. These coefficients and their second differences are tabulated at interval 0.001 correct to ten decimal places. There are supplementary tables containing exact values of these and further coefficients and their differences, at larger tabular intervals, and also a table for use in interpolating near the beginning and end of a series of tabulated values when differences are not provided. A short but adequate introduction summarizes the various formulæ and explains the use of the tables for direct and inverse interpolation and for sub-tabulation.

To the reviewer the symbolism, like much of that adopted in the calculus of finite differences, seems not entirely happy. The fraction of the interval

between two tabular values is denoted by θ , and $1-\theta$ by ϕ . The coefficients of the n th central differences for successive tabulated values u_0 and u_1 are associated with ϕ and θ respectively. The usual symbols for these coefficients, $\epsilon_n(\phi)$ and $\epsilon_n(\theta)$, avoidably insisting on the retention of two arguments for one interpolate, are made even more confusing in the Introduction by dropping the brackets "for the sake of brevity." In the tables, the argument is θ both for $\epsilon(\theta)$ and $\epsilon(\phi)$, and ϕ is tabulated at the right-hand side of the page for each value of θ . An immediate, but incorrect, implication of the setting out is, for example, that $\epsilon(\theta)$ for $\theta=0.1$ is equal to $\epsilon(\phi)$ for $\phi=0.9$. The symbolism and setting out may be compared with that in the shorter Everett coefficient tables in the Nautical Almanac, which, though not perfect, seems much more immediately indicative of how the coefficients are to be used.

These points might perhaps be considered in any future issue of these tables, particularly with the casual user of the tables in mind. The points are perhaps trivial in connexion with work involving continuous use of the tables, for care would be taken to develop a routine adapted simply to the arrangement of the relevant numbers. The tables are admirably printed, and they are sufficiently comprehensive to meet almost any imaginable computing need in the field which they cover. With a little care they could be used effectively for incidental interpolation, and for systematic sub-tabulation work they will continue to be of the greatest value.

E. C. S.

*[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]*

LII. *The Properties of Fluids.*

By S. C. BRADFORD, D.Sc.*

[Received January 22, 1943.]

ABSTRACT.

In continuation of previous work on the properties of liquids, as systems of attracting molecules, Maxwell's kinetic theory is extended to the properties of associated liquids. Maxwell's law of the distribution of molecular velocities,

$$n_u = \frac{n}{\alpha \sqrt{\pi}} e^{-\frac{u^2}{\alpha^2}} du,$$

contains an undetermined coefficient, α , the most probable speed of the molecules. The equation is adapted to the properties of gases by defining temperature as equal to the mean kinetic energy of a molecule of an ideal gas, *i. e.* by putting

$$\frac{1}{2}mC^2 \equiv \Theta,$$

where C is the mean kinetic energy velocity, which Maxwell showed to be equal to $\sqrt{\frac{3}{2}}\alpha$.

In a liquid, whose attracting molecules are mostly less than a molecular diameter apart, their average velocity is necessarily greater than in an ideal gas, where, by analogy with hydrogen, the molecules are three thousand times as far apart. By putting the most probable speed, α , equal to λP , where P is the most probable speed in an ideal gas, and allowing for the greater molecular volume of an associated liquid, Maxwell's law gives inevitably

$$A(p+K)(v-b) = R\lambda^2 T$$

as the general equation of state of a fluid. Since p is negligible compared with K , the cohesion of the liquid, we have at once

$$\lambda^2 = \frac{AK(v-b)}{RT},$$

from which λ is known, because K is given by Edser's law of force, and A can be calculated from the latent heat or viscosity, and b can be determined from any other property of a liquid.

With the aid of this equation, the latent heat, vapour pressure and viscosity of associated liquids are determined by straightforward Newtonian dynamics, without the introduction of any arbitrary constants.

* Communicated by the Author.

The method is applied also to the thermal conductivity and viscosity of gases. The resulting formulæ give values accurate to two significant figures. So we see that Edser's law of molecular attraction determines the properties of liquids, as the law of gravitation rules the motions of the heavenly bodies and Newton's principle of strict causality extends to both.

§ 1. INTRODUCTION.

IN previous papers it has been shown that the observed properties of fluids can be calculated from Newton's dynamics and Maxwell's kinetic theory adapted to the case of attracting molecules, by using Edser's inverse eighth power law of force. The marked influence of Association, due to the attraction of the molecules, has been pointed out. In this paper the exact effects of Association are determined. And, when these effects are included, it is found that the calculated properties of liquids agree with the observed values to two significant figures. The method is extended also to latent heat. Thus, molecular attraction provides a complete dynamical theory of the liquid state.

In Maxwell's expression for the number of molecules that have the components of their velocities parallel to the X-direction between u and $u+du$, *i. e.*

$$n_u = \frac{n}{\alpha \sqrt{\pi}} e^{-\frac{u^2}{\alpha^2}} du, \quad (1.1)$$

α is an arbitrary constant, whose value can be determined only from practical circumstances. A perfectly general value of α is determined from Maxwell's law as follows:

That the particles of liquids attract one another, is a fact which is demonstrated by their cohesion, surface tension, latent heat and many other properties. Whatever be the law of molecular attraction, there is a pressure, K , across any plane in the interior of a fluid, which is due to the attraction of the molecules on one side of the plane for those on the other side. This pressure is called the Cohesion of the fluid. Together with the pressure, p , upon the walls of the containing vessel and the free surface of the fluid, this pressure is balanced by the expansion pressure, p_e , due to the heat motion of the fluid particles moving across the plane. This expansion pressure keeps the molecules apart. Thus

$$p_e = p + K. \quad (1.2)$$

According to Maxwell's law, the number of molecules in unit volume of the fluid that have their velocities normal to the plane between u and $u+du$ is given by (1.1). The number that strike unit area of the plane per second is this number multiplied by u , and the partial pressure across the plane exerted by these molecules is their number multiplied by their change of momentum, $2mu$. This pressure is, therefore,

$$\frac{2mn}{\sqrt{\pi} \alpha} e^{-\frac{u^2}{\alpha^2}} u^2 du.$$

The total pressure, p_e , is obtained by integration from 0 to ∞ . This gives

$$p_e = \frac{1}{2} mn \alpha^2, \quad (1.3)$$

or,

$$p + K = \frac{1}{2} mn \alpha^2. \quad (1.4)$$

As we have to find whether the value of α is different in a liquid from its value in an ideal gas, put

$$\alpha \equiv \lambda P, \quad (1.5)$$

where P is the value of α , or the most probable speed of the molecules, in an ideal gas.

In the previous paper we defined the temperature, Θ , of the fluid as the mean kinetic energy of a particle of an ideal gas in thermal equilibrium with the fluid*. Or,

$$\Theta \equiv \frac{1}{2} m C^2, \quad (1.6)$$

where C is the mean kinetic energy velocity in a perfect gas at that temperature. And, as Maxwell has shown that

$$P^2 = \frac{2}{3} C^2, \quad (1.7)$$

we get from (1.4)

$$p + K = \frac{2}{3} n \lambda^2 \Theta, \quad (1.8)$$

Now if v is the volume of one gram mol. of the fluid, van der Waals has deduced that

$$n = \frac{N}{v-b}, \quad (1.9)$$

where

$$N = 6.06 \times 10^{23} \quad (1.10)$$

and

$$b = \frac{2}{3} \pi \sigma^3 N = 1.30 \times 10^{24} \sigma^3, \quad (1.11)$$

σ being the average of the nearest distances of approach of the centres of two molecules in a great many collisions.

So, from (1.8) and (1.9) we get

$$p + K = \frac{2}{3} \frac{N}{v-b} \lambda^2 \Theta.$$

Transposing and introducing the association factor, A , to allow for the fact that we have to take a larger volume of fluid to comprise N molecules, we obtain $A(v-b)(p+K) = \frac{2}{3} N \lambda^2 \Theta$,

in which, by putting for convenience

$$\frac{2}{3} N \equiv P, \quad (1.12)$$

we deduce the general equation of state of a real fluid as

$$A(p+K)(v-b) = P \lambda^2 \Theta. \quad (1.13)$$

It should be noticed that the factor λ^2 is a mere numerical factor, whose value has now to be determined. It is a multiplier of the temperature Θ and expresses the increase, if any, in the average kinetic energy of a molecule beyond that of a molecule of the same mass in

* *Cp. Jeans, 'Dynamical Theory of Gases,' p. 108 (1904).*

a perfect gas. It is not a multiplier of the constant P , which is always $\frac{2}{3}N$. Although it is convenient to represent it by two letters, the expression $\lambda^2\Theta$ is a single quantity, *i. e.* kinetic energy.

By a suitable change in the numerical value of P , we may write the equation (1.13) in Centigrade form as

$$A(p+K)(v-b)=\lambda^2RT. \quad (1.14)$$

This is the general equation of state of a fluid. It is the inevitable consequence of Maxwell's law as expressed in equation (1.1). It might

TABLE I.
Values of λ for liquids.

Substance.	T.	$\lambda_l=1.097 \left\{ \frac{AK_l(v-b)}{T} \right\}^{\frac{1}{2}} \times 10^{-4}.$
Acetone	329°	1.19
	298°	1.23
Ether	307.6°	1.20
	273°	1.35
<i>n</i> -Hexane.....	341°	1.21
	273°	1.45
Benzene	353.5°	1.37
	298°	1.50
Chloroform	333°	1.31
	298°	1.40
Carbon tetrachloride ..	394.4°	1.45
	298°	1.77

be described as Maxwell's equation of a real fluid. In the sequel this equation of state is deduced independently from Clausius' virial, and a further proof that $\lambda > 1$ is deduced directly from the law of the conservation of energy.

If we put $A=1, K=0, b=0, \lambda=1, \quad (1.15)$

as in an ideal gas, we obtain $pv=RT$,

the well-known equation of Boyle's law. We see that this equation depends on the truth of equations (1.15). That is to say, it is the equation of an ideal gas only. It cannot apply to liquids, because the equations (1.15) are certainly not all true for them.

To obtain the numerical values of λ for liquids, we have merely to

notice that for these the value of the atmospheric pressure, p , is negligible in comparison with that of the cohesion, K . So we derive

$$\text{from (1.14)} \quad \lambda^2 = \frac{AK(v-b)}{RT} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.16)$$

$$\text{or,} \quad \lambda = 1.097 \left\{ \frac{AK(v-b)}{T} \right\}^{\frac{1}{2}} \times 10^{-4}. \quad . \quad . \quad . \quad . \quad . \quad (1.17)$$

In previous papers, we have shown that K can be calculated from the observed surface tension of the liquid by Edser's method, modified to allow for the change of density through the surface of the liquid. Methods of determining A from several properties are developed in this paper. Consequently all the factors on the R.H.S. of (1.17) are known and the numerical values of λ can be calculated. For liquids examined in this paper these values are given in Table I.

From this table we see that in liquids λ is certainly not equal to unity, and consequently, the most probable speed of the molecules is greater than in an ideal gas at the same temperature. If we realize that the average distance travelled by a molecule of hydrogen between two collisions is about 2,000 times the diameter of a molecule, whereas the average distance apart of two molecules of a liquid is about half the diameter of a molecule, it is clear that the effect of molecular attraction in a liquid is altogether greater. So that, in gases, the effect of the attraction is negligible, except under very great compression, or at low temperatures, whereas, in liquids, molecules are always under the influence of great attraction, except when passing through points where the resultant of the molecular forces momentarily vanishes. Consequently, in gases λ is equal to unity and the equations derived from the theory of motion in generalized space hold; but in liquids, as Jeans points out at the end of his 'Dynamical Theory of Gases,' p. 329 (1904), "the analysis ceases to apply, because the assumption we made . . . that the [number of] molecular clusters is small, is now invalidated." In fact the whole liquid is one liquid cluster, which is beginning to contain new aggregates in the form of incipient solid particles. In the present theory, the analysis is adapted to this fact.

To fix our ideas, the actual values of the mean path of the molecules in some of the liquids considered are given in Table II.

It should be noted that these values are the same for all properties considered.

Maxwell deduced his law of the equipartition of energy between gases at the same temperature, from his equation in which $\lambda=1$. It will be seen that his law is really a statement of the equalization of temperature. And in this form, it holds for all substances. In liquids, the potential energy of the closely adjacent attracting molecules is different. The theory of molecular attraction in no way contradicts the Maxwell's law of equipartition in gases, it extends this law to liquids, and, it can be expected, to solids as well.

By means of equations (1.16) or (1.17), λ can be eliminated from the equations expressing the properties of liquids, giving formulæ in which all the factors can be determined. The theory is more than a mere mathematical ideal, it constitutes a physical mechanism, by which the hidden phenomena of molecular attraction and motion can be followed in detail and their effects evaluated.

In the last paper⁽¹²⁾, a number of these effects were examined. It was shown that all gases and liquids are associated in greater or less degree. According to Maxwell's law of the distribution of velocities, at any instant, in any real fluid, always some molecules approach on

TABLE II.

The mean paths of molecules in liquids.

Substance.	Absolute temperature. T.	Molecular diameter. $\sigma \times 10^8$.	Mean path $\frac{1}{\sqrt{2} \pi n \sigma^2} \times 10^8$.
Ether.....	273°	4.10	2.19
	307°	4.12	2.30
<i>n</i> -Hexane	273°	4.45	2.49
	341°	4.51	2.55
Benzene	298.1°	3.96	2.09
	353.5°	4.00	2.20
Chloroform	298°	3.81	2.04
	333.2°	3.84	2.11
Carbon tetrachloride.....	298°	4.06	2.38
	349.4°	4.10	2.43

another with relative velocities less than the critical velocity, just sufficient to overcome their mutual attraction. Such molecules will remain in association until, in the molecular chaos, their relative velocities become great enough to separate them. Even the noble gases are appreciably associated at very low temperatures, or at very high pressures.

Besides the change in the most probable speed and molecular association, a third effect of molecular attraction is the rapid, but continuous, change of density that occurs in passing through the interface between any two different media or phases. As shown in the last paper, this phenomenon requires the modification of van der Waals' collision factor,

$\frac{v}{v-b}$, which represents the result of the transfer of momentum through the distance of a molecular diameter at each collision. For a liquid in

contact with its vapour, the factor became $\frac{v}{v-\frac{b}{2}}$. At solid-liquid and

solid-gas surfaces, molecular attraction causes an increase of density of the liquid, or the condensation of the gas, as well as an increase in the speeds of the molecules. The presence of this so-called "adsorbed" layer and this increased speed, affects the measurement of almost every physical property, because it is impossible to experiment with a fluid without isolating it, thus creating a surface.

A fourth effect, closely related, is the change in the cohesion, resulting in the alteration of the value of the mean square velocity, λC , in passing from one medium to another, which introduces a gain or loss of potential energy.

As we have seen, the theory gives a mental picture of the internal phenomena of gases and liquids. Molecules behave as smooth elastic attracting spheres. Although the actual conformation of a molecule is not spherical, the physical properties we observe are average effects. Collisions take place so rapidly that we are unable to observe what happens during a single encounter. We see the effect of a great many collisions, in which the average of the nearest distances of approach of two molecules is σ . So that we can say that, on the average, collision between two molecules takes place when their centres come to a mean distance σ apart, and, as Edser pointed out, the molecules attract one another, even at the instant of collision.

The main principles of the theory are independent of any particular law of force. The formulæ can be adapted to any law of force that may be found to hold. But Edser's law of the inverse eighth power

$$f = \frac{\mu}{r^8} \quad \dots \quad (1.18)$$

gives values for the properties of liquids exact to two significant figures.

From this law we calculate the cohesion of a liquid

$$K = \frac{\pi \mu n^2}{6 \sigma^4} = \frac{\pi \mu N^2}{6 \sigma^4 v^2} \quad \dots \quad (1.19)$$

and the surface tension

$$S = \frac{\pi \mu n^2}{24 \sigma^3} \cdot \frac{v - \frac{b}{2}}{v} = \frac{\pi \mu N^2 \left(v - \frac{b}{2} \right)}{24 \sigma^3 v^3}, \quad \dots \quad (1.20)$$

so that we have a means of calculating the cohesion from the observed surface tension by the formula

$$K = \frac{4S}{\sigma} \cdot \frac{v}{v - \frac{b}{2}} = \frac{4 \cdot 331 S}{b^{1/3}} \cdot \frac{v}{v - \frac{b}{2}} \times 10^8. \quad \dots \quad (1.21)$$

We are now in a position to calculate the work done by a molecule in passing from one phase, a , into another, b . This comprises (1) the

work done against the attractive forces of medium a and by those of medium b ; and (2) the work done by the change in potential energy. For a molecule of a liquid escaping into its vapour, the total work done is

$$[W_e]_l = \frac{1}{N}(K_l v_l + P\Theta). \quad (1.22)$$

From this formula, the latent heat of an unassociated liquid is evaluated as

$$L = \frac{1.98}{M}(5.206\chi_l v_l + T), \quad (1.23)$$

where χ is the cohesion in special units. The analysis is then extended to associated liquids.

The vapour pressure of the liquid is formed of the particles that have sufficient energy, normal to the surface of the liquid, to do the work involved in escape. The vapour pressure is determined by equating the number of these particles escaping from unit area of the surface to the number striking that surface from the vapour. This gives, for an unassociated liquid,

$$p_{mm} = \left(\frac{2\chi_l T}{v_l - b}\right)^{\frac{1}{2}} e^{-\frac{v_l + 0.192\frac{T}{\chi_l}}{v_l - b}} \times 10^5. \quad (1.24)$$

Preliminary treatment is given to the thermal conductivity of gases. This illustrates the preponderating effect of the attraction of the molecules comprising the solid containing vessel on those of the gas. The formula for the thermal conductivity of gases is

$$K = 2.12 \times 10^{-5} M \frac{T}{M^{\frac{1}{2}} b^{\frac{2}{3}}} \lambda_0^3 \frac{v_0}{v_0 - b}, \quad (1.25)$$

where M is the number of modes of motion of the molecule and v_0 is the molecular volume of the compressed layer of gas at the solid surface.

The only gas considered for which this fraction, $\frac{v_0}{v_0 - b}$, has appreciable effect is hydrogen. λ_0 is the value of λ at the solid surface as increased by the attraction of the solid particles. This effect is appreciable for all gases and is practically constant with change of temperature. It is the same for both the thermal conductivity and viscosity of gases. This analysis suggests some alterations in the equations for the viscosities of gases and liquids, *i. e.*

$$\eta_g = 1.93 \times 10^{-5} \frac{M^{\frac{1}{2}} T^{\frac{1}{2}}}{b^{\frac{2}{3}}} \lambda_0^{\frac{1}{2}} \frac{v_0 - \frac{b}{2}}{v_0 - b}, \quad (1.26)$$

and

$$\eta_l = 2.30 \times 10^{-5} \left(\frac{\chi M}{v_l - b}\right)^{\frac{1}{2}} \frac{v_l + 3b}{b^{\frac{3}{2}}}. \quad (1.27)$$

For gases, the method of finding v_0 and λ_0 is indicated and the values needed are tabulated. In liquids, the equation does not contain such values, and all quantities needed can be found from the published

tables of physical constants. The effect of association on all these properties has been determined.

It will be seen that all these equations, derived directly from Newton's laws of motion and Maxwell's law of the distribution of molecular velocities, contain no arbitrary factors. The figures for the properties of liquids calculated from these equations agree so exactly with the observed values of these properties as to leave no doubt that they depend on molecular attraction, volume, mass and motion only.

§ 2. THE LAW OF FORCE.

The general principles developed in this series of papers depend on the fact that the molecules of substances attract one another. The basic theory is not dependent on the adoption of any particular law of force. But the observed properties of matter are the result of a definite law and we cannot proceed to calculate them without assuming such a law. As has been said, the law of force adopted is Edser's inverse eighth power law, which is found to fit the facts exactly. As Edser pointed out, however, molecules are electrically neutral, they attract in some orientations and repel in others. Edser suggested that the resultant attractive effect is due to the longer duration of the attractive encounters. A little consideration shows that the preponderance of the attractive effect is the result of the fact that, in attractive orientations, the molecules are brought nearer together, whereas, in repulsive orientations, the molecules are pushed away from each other and, indeed, unless the repelling molecules have a sufficient relative velocity they will not even collide. But, all approaching molecules will come into actual collision when their orientations are such that they mutually attract one another. It should not be difficult to deduce a general law of force, involving both attraction and repulsion, from this idea. But, in this paper, Edser's simple law is sufficient.

Thus, in dealing with the properties of liquids, we must bear in mind that they are average effects. We cannot measure what happens to two molecules in an instant of time. We observe only the result of a great many individual collisions, during a short period of time. And these effects are found to be constant over any period. Molecules are not spherical in shape. The average nearest distance of approach of the molecules, σ , is not usually the distance apart of two centres in a particular collision, but the average result of many collisions, which is found to be constant. So, if we take any space element dv , the molecules with their centres within dv will be changing continually in number, mutual attraction and nearest distances of approach. But, during a sufficient small interval of time, which can be smaller than we can observe, any point within dv will have been occupied by a molecular centre as often as any other point, and the average density, force and nearest distance of approach will be the same as at any other point. Thus we are entitled to regard as constant, n , the number of molecules

per unit volume, μ , the apparent coefficient of attraction of two molecules and σ , the nearest distance of approach of two molecular centres.

§ 3. THE ASSOCIATION, COHESION AND SURFACE TENSION OF ASSOCIATED LIQUIDS.

We have seen that Maxwell's law of the distribution of molecular velocities applies to liquids as well as to gases, and that the most probable speed, α , is greater in a liquid because the attraction of its particles increases their velocities. To express this difference, it is convenient to write Maxwell's law in the form

$$n_u = \frac{n}{\sqrt{\pi} \lambda \mathbf{P}} e^{-\frac{u^2}{\lambda^2 \mathbf{P}^2}} du, \quad (3.1)$$

where n_u is the number of molecules per unit volume that have velocities between u and $u+du$, n is the number of molecules in unit volume of the liquid, and λ is a number expressing the ratio of the most probable speed, α , in a liquid to the most probable speed, \mathbf{P} , in an ideal gas of the same molecular weight at the same temperature.

According to Maxwell's law, at any instant there will always be some pairs of molecules that approach each other with relative velocities less than the critical velocity, which is sufficient to separate them against their mutual attraction. These pairs of molecules will remain associated until, in the molecular chaos, their relative velocities increase beyond the critical value. This applies to gases as well as liquids, although in gases association forms incipient liquid particles free to revolve mutually, while in liquids, association produces particles akin to the solid form and fixed in the position of optimum attraction. Therefore, unit volume of such fluids will contain both single and double molecules, or even particles more associated. Let n_1 and n_2 be the numbers of single and double molecules respectively per c.c. If m is the mass of a single molecule, the mass of a double molecule is $2m$, and the average mass m_{12} of the n_{12} particles per unit volume is

$$m_{12} = \frac{n_1 m_1 + 2n_2 m_1}{n_{12}}, \quad (3.2)$$

Multiplying by N , the number of particles in the molecular volume, and putting

$$M_{12} \equiv AM, \quad (3.3)$$

where A is the association factor and M is the molecular weight, we get

$$AM = M \frac{n_1 + 2n_2}{n_{12}},$$

or

$$A = \frac{n_1 + 2n_2}{n_{12}}, \quad (3.4)$$

This gives

$$n_1 = n_{12}(2-A), \quad (3.5)$$

$$n_2 = n_{12}(A-1), \quad (3.6)$$

equations we shall need in the sequel.

The relation between the cohesion and surface tension of an unassociated liquid was found in the last paper as

$$K = \frac{4 \cdot 331 S v}{b^{1/3} \left(v - \frac{b}{2} \right)} \times 10^8. \quad (3.7)$$

In an associated liquid, the actual volume of the double molecules is necessarily about twice that of the single molecules. We should expect a slight contraction in the formation of incipient solid particles, but we cannot evaluate this at present and its amount is practically negligible. It follows that

$$b_{22} = 2b_{11} \quad \text{and} \quad b_{12} = Ab_{11}. \quad (3.8)$$

So we get

$$K_{12} = \frac{4 \cdot 331 S_{12}}{A^{1/3} b_{11}^{1/3}} \cdot \frac{v}{v - \frac{b}{2}} \times 10^8, \quad (3.9)$$

where S_{12} is the observed surface tension of the associated liquid. This equation leads to the approximate evaluation of the attraction coefficient of the double molecules. We saw in the previous paper ⁽¹²⁾ that the cohesion of an associated liquid is expressed by the formula

$$K_{12} = \frac{\pi}{6} \left(\frac{n_1 \mu_1^{\frac{1}{2}}}{\sigma_1^2} + \frac{n_2 \mu_2^{\frac{1}{2}}}{\sigma_2^2} \right)^2, \quad (3.10)$$

and from (3.8) we have

$$\sigma_2 = 2^{\frac{1}{2}} \sigma_1, \quad (3.11)$$

so that, if $\mu_2 = x\mu_1$,

$$K_{12} = \frac{\pi \mu_1}{6 \sigma_1^4} \left(n_1 + \frac{n_2 x^{\frac{1}{2}}}{2^{2/3}} \right)^2. \quad (3.12)$$

Moreover, from the previous paper and, neglecting the collision factor,

$\frac{v}{v - \frac{b}{2}}$, which is not affected by association,

$$S_{12} = \frac{\pi}{24} \left(\frac{n_1 \mu_1^{\frac{1}{2}}}{\sigma_1^{3/2}} + \frac{n_2 \mu_2^{\frac{1}{2}}}{\sigma_2^{3/2}} \right)^2 \quad (3.13)$$

$$= \frac{\pi \mu_1}{24 \sigma_1^3} \left(n_1 + \frac{n_2 x^{\frac{1}{2}}}{2^{\frac{1}{2}}} \right)^2. \quad (3.14)$$

If we put

$$x = 2, \quad (3.15)$$

(3.14) gives

$$\begin{aligned} S_{12} &= \frac{\pi \mu_1}{24 \sigma_1^3} (n_1 + n_2)^2 \\ &= \frac{\pi \mu_1 n_{12}^2}{24 \sigma_1^3}, \end{aligned} \quad (3.16)$$

and (3.12) becomes

$$K_{12} = \frac{\pi \mu_1}{6 \sigma_1^4} \left(n_1 + \frac{n_2}{2^{1/6}} \right)^2. \quad (3.17)$$

Now, it is obvious that when the liquid comprises only double molecules, n_1 being zero, $A=2$, and $n_2=n_{12}$, we get

$$n_1 + \frac{n_2}{2^{1/6}} = \frac{n_{12}}{A^{1/6}},$$

which can be verified for other cases. So we may take the equation

$$n_1 + \frac{n_2}{2^{1/n}} = \frac{n_{12}}{A^{1/n}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.18)$$

as being, at least, approximately two. This very powerful approximate expression will be used in the sequel.

From this we get from (3.17)

$$\begin{aligned} K_{12} &= \frac{\pi\mu_1}{6\sigma_1^4} \cdot \frac{n_{12}^2}{A^{1/3}} \\ &= \frac{4S_{12}}{A^{1/3}\sigma} \cdot \frac{v}{v-\frac{b}{2}} \\ &= \frac{4 \cdot 331 S_{12}}{A^{1/3} b^{1/3}} \cdot \frac{v}{v-\frac{b}{2}} \times 10^8, \quad . \quad . \quad . \quad . \quad . \quad (3.19) \end{aligned}$$

which corresponds exactly to (3.9). We find, therefore, that the attraction coefficient between two double molecules, μ_2 , is just twice the value of that expressing the force between two single molecules, *i. e.*

$$\mu_2 = 2\mu_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.20)$$

This very important relation enables us to calculate the effect of double molecules in associated liquids.

Before leaving the matter of the cohesion and surface tension of associated liquids we must bear in mind certain facts :

Because
$$n_{12} = \frac{n}{A}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.21)$$

where n is the number of single molecular units in one c.c. of liquid and

$$n = \frac{N}{v}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.22)$$

v being the volume of liquid containing the molecular weight in grams of the unassociated molecule, or $v = \frac{M}{\rho}$, $. \quad . \quad . \quad . \quad . \quad . \quad (3.23)$

where ρ is the density of the liquid,

$$\therefore v_{12} = Av. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.24)$$

So we have from (3.19)
$$K_{12} = \frac{\pi\mu_1}{6\sigma_1^4} \frac{n^2}{A^{7/3}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.25)$$

$$= \frac{K_{11}}{A^{7/3}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.26)$$

and from (3.16)
$$S_{12} = \frac{\pi\mu_1 n^2}{24\sigma_1^3 A^2} \cdot \frac{v-\frac{b}{2}}{v} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.27)$$

$$= \frac{S_{11}}{A^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.28)$$

These latter equations are necessary for calculating the value of the attraction coefficients of associated liquids. But, in actual practice, we do not know the value of the association factor in advance and we begin by measuring the surface tension of the liquid and calculating the

$$\text{cohesion from the formula } K_{1l} = \frac{4 \cdot 331 S_{12l}}{b^{1/3}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.29)$$

in which the subscript l is used to denote the liquid under observation and the combination $1l$ is a symbol to express the fact that the cohesion is the value of the R.H.S. of equation (3.29). This fictitious value is simply K_{12l} , without the $A^{1/3}$, or

$$K_{1l} = A^{1/3} K_{12l}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.30)$$

This value of the cohesion will be needed in the calculations that follow. It will be observed that these relationships are perfectly simple, but must be carefully borne in mind in evaluating the properties of associated liquids.

§ 4. THE POTENTIAL ENERGY OF A LIQUID.

Imagine an ideal liquid comprising molecules like those of Maxwell's perfect gas, but having the additional property of mutual attraction. It does not matter what law governs the attraction, but we may bear in mind Edser's inverse eighth power law. Take a vacuous vessel impervious to heat by any process and introduce some of the liquid, closing the vessel. As particles of the liquid approach its surface with sufficient kinetic energy of translation to overcome the attraction of the bulk of the liquid, these particles will escape into the vacuum and form a vapour. This will increase in density until as many particles from the vapour strike the surface of the liquid per second as evaporate from the liquid. Thereafter, the concentration of the vapour will remain constant. To fix our ideas, let us suppose that we have now an equal number of molecules in both states, liquid and vapour. We may take this number as N , the number of molecules in a gram mol. This system of attracting molecules constitutes a conservative system, because the work done by an escaping molecule or on a returning molecule depends only on the attractions of the molecules and not on their paths. Now the average distance apart of the molecules in the two states will be very different. By analogy with hydrogen, for example, the surfaces of the molecules in the liquid will be an average of about half a molecular diameter from each other, whereas those in the vapour will be some three thousand diameters apart. The attraction between the molecules in the liquid state, so close together, will be very considerable, whereas that in the vapour state will be nothing to measure. Thus we may take the vapour as the standard state of no potential energy, while the liquid will have great negative potential energy. Thus, if Φ_g and Φ_l are the potential energies in the vapour and the liquid,

$$\Phi_g > \Phi_l. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.1)$$

Now, if E_g and E_l are the corresponding kinetic energies of translation,

$$\Phi_l + E_l = \Phi_g + E_g \quad (4.2)$$

by the principle of the Conservation of Energy.

$$\text{Let us put} \quad E_l \equiv \lambda^2 E_g, \quad (4.3)$$

where λ^2 is a numerical factor.

$$\text{Then} \quad \Phi_l + \lambda^2 E_g = \Phi_g + E_g \quad (4.4)$$

$$\text{or} \quad \Phi_g - \Phi_l = (\lambda^2 - 1) E_g, \quad (4.5)$$

$$\text{and by (4.1)} \quad \lambda^2 > 1. \quad (4.6)$$

This is an independent proof of the experimental fact shown in Table I., that in liquids, λ has a value greater than unity.

Now, as Dieterici showed, the total kinetic energy of translation which is transferred to the vapour when a gram mol. of liquid is vaporized is, using λP for Dieterici's β and P for his α ,

$$E_l = \frac{1}{2} N m \lambda^2 P^2 = P \lambda^2 \Theta, \quad (4.7)$$

and that brought back by a gram mol. of the vapour is

$$E_g = \frac{1}{2} N m P^2 = P \Theta. \quad (4.8)$$

Thus, we have, per gram mol., putting $\Phi_g = 0$ in the standard state,

$$\Phi_l = -(\lambda^2 - 1) P \Theta. \quad (4.9)$$

This is the potential energy of a gram mol. of an ideal liquid.

§ 5. THE WORK DONE BY A MOLECULE IN PASSING FROM ONE FLUID PHASE INTO ANOTHER.

In § 11 of the previous paper ⁽¹²⁾ we determined the work done by a molecule in escaping from a homogeneous fluid through the surface layers of continuously decreasing density into its vapour. As we shall have to deal with fluids containing two kinds of molecules, *e. g.* single and associated molecules, or those of a solute and its solvent, it is necessary to extend the investigation to cover this case.

Let the two media, which may be a liquid and a gas or two liquids, be separated by a horizontal plane, $x=0$, as in § 11. Denote the two media by the letters **a** and **b***, of which **b** may be the denser and lower. In the condition of thermal equilibrium, each medium will contain a definite concentration of each of the two kinds of molecules. Let these be denoted by the subscripts 1 and 2.

(i) *The Work Against the Forces of Attraction.*

In considering the work done by a molecule in moving from **b** to **a**, we have to allow for van der Waals' collision factor, $\frac{v-b}{v}$, which represents

* When the letters denoting the media occur as subscripts these are printed in ordinary italic type.

the effect of the transfer of momentum through a molecular diameter in collision with another molecule. So we have

$$\begin{aligned}\frac{v-b}{v} &= 1 - \frac{b}{v} \\ &= 1 - \frac{bf(\xi)}{N}, \quad (5.1)\end{aligned}$$

where $v = \frac{N}{n}$

and $n \equiv f(\xi), \quad (5.2)$

ξ being the x co-ordinate of the layer of fluid and n the number of particles per unit volume.

In the previous paper, we found that the work done by a molecule in traversing the distance from the interior of medium **b** to that of

medium **a** is $W_f = -\frac{\pi\mu}{3\sigma^4} \left[\int_{+\infty}^{\infty} f'(\xi) d\xi - \frac{b}{N} \int_{+\infty}^{\infty} f(\xi) f'(\xi) d\xi \right]. \quad . . (5.3)$

In a mixture we have to consider the effect of the attraction of molecules of both kinds on a particle, which we will suppose is of kind 2. It is easy to show that, if $f(\xi_2)$ is the concentration of the molecules of kind 2, the work done against the attraction of all the molecules of kind 2 in medium **b**, and by the attraction of similar molecules in medium **a**, is

$$\begin{aligned}W_{f22} &= -\frac{\pi\mu_2}{3\sigma_2^4} \left[\int_{+\infty}^{\infty} f'(\xi_2) d\xi_2 - \frac{b_2}{N} \int_{+\infty}^{\infty} f(\xi_2) f'(\xi_2) d\xi_2 \right] \\ &= -\frac{\pi\mu_2}{3\sigma_2^4} \left\{ \left(n_{2b} - \frac{b_2}{2N} n_{2b}^2 \right) - \left(n_{2a} - \frac{b_2}{2N} n_{2a}^2 \right) \right\} \\ &= -\frac{\pi\mu_2}{3\sigma_2^4} \left\{ n_{2b} \left(1 - \frac{b_2}{2N} n_{2b} \right) - n_{2a} \left(1 - \frac{b_2}{2N} n_{2a} \right) \right\} \quad . . (5.4)\end{aligned}$$

Similarly, $W_{f21} = -\frac{\pi\mu_{21}}{3\sigma_{21}^4} \left[\int_{+\infty}^{\infty} f'(\xi_1) d\xi_{11} - \frac{b_{21}}{N} \int_{+\infty}^{\infty} f(\xi_1) f'(\xi_1) d\xi_1 \right]$

$$= -\frac{\pi\mu_2^{\frac{1}{2}}\mu_1^{\frac{1}{2}}}{3\sigma_1^2\sigma_2^2} \left\{ n_{1b} \left(1 - \frac{b_{21}n_{1b}}{2N} \right) - n_{1a} \left(1 - \frac{b_{21}n_{1a}}{2N} \right) \right\} \quad . (5.5)$$

So that the total work done by the molecule of kind 2 against the attraction of all the molecules of medium **b**, and by the attraction of all the molecules of medium **a** on the molecule is

$$\begin{aligned}[W_{f2}]_b &= -\frac{\pi\mu_2^{\frac{1}{2}}}{3\sigma_2^2} \left[\frac{\mu_2^{\frac{1}{2}}}{\sigma_2^2} \left\{ n_{2b} \left(1 - \frac{b_2 n_{2b}}{2N} \right) - n_{2a} \left(1 - \frac{b_2 n_{2a}}{2N} \right) \right\} \right. \\ &\quad \left. + \frac{\mu_1^{\frac{1}{2}}}{\sigma_1^2} \left\{ n_{1b} \left(1 - \frac{b_{21} n_{1b}}{2N} \right) - n_{1a} \left(1 - \frac{b_{21} n_{1a}}{2N} \right) \right\} \right] \quad . (5.6)\end{aligned}$$

Now

$$\frac{b_2 n_2}{2N} = \frac{b_2}{2v_2},$$

and

$$b_b = \frac{n_{1b}b_1 + n_{2b}b_2}{n_b} \text{ approximately.}$$

So that

$$\frac{b_b}{v_b} = \frac{b_1}{v_1} + \frac{b_2}{v_2}.$$

In the case of an associated liquid having single and double molecules, it should be noted that $\frac{b_b}{v_b}$ is the same value as $\frac{b}{v}$.

To continue, if b_{21} is not greatly different from b_1 , nor $\mu_1^{\frac{1}{2}}$ from $\mu_2^{\frac{1}{2}}$, all involving square roots, we may re-write (5.6)

$$\begin{aligned} [W_{f2}]_b^a = & -\frac{\pi\mu_2^{\frac{1}{2}}}{3\sigma_2^2} \left(1 - \frac{b_b}{2v_b}\right) \left\{ \frac{\mu_2^{\frac{1}{2}}n_{2b}}{\sigma_2^2} + \frac{\mu_1^{\frac{1}{2}}n_{1b}}{\sigma_1^2} \right\} \\ & + \frac{\pi\mu_2^{\frac{1}{2}}}{3\sigma_2^2} \left(1 - \frac{b_a}{2v_a}\right) \left\{ \frac{\mu_2^{\frac{1}{2}}n_{2a}}{\sigma_2^2} + \frac{\mu_1^{\frac{1}{2}}n_{1a}}{\sigma_1^2} \right\}. \quad (5.7) \end{aligned}$$

In this expression $\frac{\pi\mu_2^{\frac{1}{2}}}{3\sigma_2^2} \left(1 - \frac{b_b}{2v_b}\right) = \frac{2K_{2b}^{\frac{1}{2}}}{N} \left(v_b - \frac{b_b}{2}\right),$

and similarly for the factors with subscript a , while the expressions in curly brackets are proportional to $K_{21b}^{\frac{1}{2}}$, etc. Thus we get

$$[W_{f2}]_b^a = - \left\{ \frac{2K_{2b}^{\frac{1}{2}}K_{21b}^{\frac{1}{2}}}{N} \left(v_b - \frac{b_b}{2}\right) - \frac{2K_{2a}^{\frac{1}{2}}K_{21a}^{\frac{1}{2}}}{N} \left(v_a - \frac{b_a}{2}\right) \right\}, \quad (5.8)$$

with the provisos mentioned above.

(ii) The Effect of the Change in Potential Energy.

As in (4.5) we have, only the kinetic energy of translation being involved,

$$\Phi_a - \Phi_b = (\lambda_b^2 - \lambda_a^2) E_a \quad (5.9)$$

$$= (\lambda_b^2 - \lambda_a^2) P\Theta. \quad (5.10)$$

To calculate λ_b and λ_a , we make use of the virial of Clausius, according to which the kinetic energy of all the molecules in a medium is expressed by the equation

$$\Sigma \frac{1}{2} m C_r^2 = -\frac{1}{2} \Sigma (xX + yY + zZ), \quad (5.11)$$

where C_r is the kinetic energy velocity of a particle of mass m , x , y , z are the co-ordinates of the particle and X , Y , Z the resolved components of all the forces acting on it.

In considering the forces of attraction between the molecules, let us suppose, at first, that they are all the same, and let the distance between a pair be r . The force of attraction between them is, by Edser's law,

$$\phi(r) = \frac{\mu}{r^8}. \quad (5.12)$$

So that

$$-\frac{1}{2} \Sigma (xX + yY + zZ) \text{ becomes } -\frac{1}{2} \Sigma r\phi(r),$$

where

$$-\frac{1}{2} r\phi(r) = -\frac{1}{2} \frac{\mu}{r^7}. \quad (5.13)$$

Now, the chance of a molecule having its centre in a shell of thickness dr at a distance r from another molecule is $\frac{4\pi r^2 dr}{v}$, and the number of pairs of molecules is $\frac{N(N-1)}{2}$ or $\frac{N^2}{2}$, neglecting 1 as compared with the extremely large number N^2 . Thus, the number of pairs of molecules having their centres within a distance r of one another is $\frac{2\pi N^2}{v} r^2 dr$.

Consequently, the virial of the attractive forces becomes

$$-\frac{1}{2}\Sigma r\phi(r) = -\int_a^\infty \frac{\pi N^2 \mu}{v} \frac{dr}{r^5} = \frac{\pi \mu N^2}{4\sigma^4 v}. \quad (5.14)$$

But we have seen that $K = \frac{\pi \mu N^2}{6\sigma^4 v^2}$,

which makes the r.h.s. of the virial equation

$$-\frac{1}{2}\Sigma r\phi(r) = \frac{3}{2}Kv. \quad (5.15)$$

Now, according to Lorenz's analysis, the virial of the forces of restitution on collision of the molecules is

$$-\frac{\pi m \sigma^3 N^2}{3v} C_r^2,$$

and we have seen that the effect of the atmospheric pressure on a liquid is negligible compared with that of its cohesion. Moreover,

$$\frac{1}{2}\Sigma m C_r^2 = \frac{1}{2}N m C_r^2. \quad (5.16)$$

Consequently, the virial equation becomes

$$\frac{1}{2}N m C_r^2 = \frac{3}{2}Kv - \frac{1}{2}m C_r^2 \cdot \frac{2}{3} \frac{\pi \sigma^3 N^2}{v},$$

or

$$3Kv = N m C_r^2 \left(1 + \frac{2}{3} \frac{\pi \sigma^3 N}{v}\right),$$

where the factor in parentheses equals $\left(1 + \frac{b}{v}\right)$, which gives

$$3Kv = N m C_r^2 \left(1 + \frac{b}{v}\right).$$

And as

$$\frac{1}{2}m C_r^2 = \lambda^2 \Theta,$$

we have approximately

$$Kv \left(1 - \frac{b}{v}\right) = \frac{2}{3}N \lambda^2 \Theta,$$

or

$$K(v-b) = P \lambda^2 \Theta, \quad (5.17)$$

which is an independent proof of the equation of state of a liquid medium, from which it follows, as before, that $\lambda > 1$.

We can now apply this method to the determination of λ_{2b} and λ_{2a} , corresponding to molecules of kind 2 in both media.

For like molecules, we have

$$-\frac{1}{2}r\phi(r) = -\frac{1}{2} \frac{\mu_2}{r^7}, \quad (5.18)$$

and the number of pairs of molecules of kind 2 at a distance between r and $r+dr$ in the volume v_b

$$= \frac{2\pi n_{2b}^2 v_b^2}{v_b} \cdot r^2 dr = 2\pi n_{2b}^2 v_b r^2 dr.$$

So that the virial of the attractive forces between molecules of kind 2

$$\text{in medium } \mathbf{b} \text{ is } -\frac{1}{2}\Sigma r(\phi)r = -\int_{\sigma_{22}}^{\infty} \pi n_{2b}^2 v_b \mu_2 \frac{dr}{r^5} = \frac{\pi \mu_2 n_{2b}^2 v_b}{4\sigma_{22}^4}. \quad (5.19)$$

Similarly for the forces between molecules of different kinds,

$$-\frac{1}{2}r\phi(r) = -\frac{1}{2} \frac{\mu_2^{\frac{1}{2}} \mu_1^{\frac{1}{2}}}{r^7},$$

and the number of pairs at a distance between r and $r+dr$ is

$$2\pi n_{2b} n_{1b} v_b r^2 dr,$$

Giving the virial in this case

$$-\frac{1}{2}\Sigma r\phi(r) = -\int_{\sigma_{21}}^{\infty} \pi n_{2b} n_{1b} v_b \mu_2^{\frac{1}{2}} \mu_1^{\frac{1}{2}} \frac{dr}{r^5} = \frac{\pi \mu_2^{\frac{1}{2}} \mu_1^{\frac{1}{2}} n_{1b} n_{2b} v_b}{4\sigma_{21}^4}. \quad (5.20)$$

Consequently, the total energy due to the forces of attraction of the molecules of kind 2 between themselves and between molecules of

$$\text{kind 1 is } \frac{\pi \mu_2^{\frac{1}{2}} n_{2b}}{4\sigma_2^2} \left(\frac{\mu_2^{\frac{1}{2}} n_{2b}}{\sigma_2^2} + \frac{\mu_1^{\frac{1}{2}} n_{1b}}{\sigma_1^2} \right) v_b.$$

Thus the total energy of all the molecules of kind 2 in medium \mathbf{b} is

$$\frac{1}{2} n_{2b} m_2 v_b \lambda_{2b}^2 C_2^2 = \frac{\pi \mu_2^{\frac{1}{2}} n_{2b}}{4\sigma_2^2} \left(\frac{\mu_2^{\frac{1}{2}} n_{2b}}{\sigma_2^2} + \frac{\mu_1^{\frac{1}{2}} n_{1b}}{\sigma_1^2} \right) v_b. \quad (5.21)$$

The energy of one molecule is obtained by dividing through by n_{2b} . So, making use of the definition of temperature,

$$\frac{1}{2} m_2 C_2^2 \equiv \Theta,$$

and introducing Lorentz's correction, we get

$$v_b \lambda_{2b}^2 \Theta = \frac{3}{2} \frac{K_{2b}^{\frac{1}{2}}}{N} K_b^{\frac{1}{2}} (v_b - b_b) v_b, \quad (5.22)$$

which gives

$$\lambda_{2b}^2 = \frac{K_{2b}^{\frac{1}{2}} K_b^{\frac{1}{2}} (v_b - b_b)}{P \Theta}, \quad (5.23)$$

as

$$P = \frac{2}{3} N.$$

Obviously, expressions for λ_{2a} , λ_{1b} and λ_{1a} are obtained by merely changing the subscripts.

Thus the change in potential energy when a molecule of kind 2 passes from medium \mathbf{b} to medium \mathbf{a} is

$$\begin{aligned} [W_{\Phi 2}]_b^a &= -\frac{2}{3} (\lambda_{2b}^2 - \lambda_{2a}^2) \Theta \\ &= -\frac{1}{N} \{ K_{2b}^{\frac{1}{2}} K_b^{\frac{1}{2}} (v_b - b_b) - K_{2a}^{\frac{1}{2}} K_a^{\frac{1}{2}} (v_a - b_a) \}. \quad (5.24) \end{aligned}$$

Consequently the total work to be done by a molecule of kind 2 in moving from medium **b** to medium **a** is

$$[W_{s2}]_b^a = [W_{f2}]_b^a + [W_{\Phi 2}]_b^a,$$

which reduces to the simple expression

$$[W_{s2}]_b^a = \frac{1}{N} (K_{2b}^{\frac{1}{2}} K_b^{\frac{1}{2}} v_b - K_{2a}^{\frac{1}{2}} K_a^{\frac{1}{2}} v_a). \quad (5.25)$$

In case the molecules are of the same kind, as when an unassociated liquid evaporates into its vapour, it is easily shown that the formula

$$\text{becomes} \quad [W_s]_l^g = \frac{1}{N} (K_l v_l + P\Theta). \quad (5.26)$$

It will be seen that the equation (5.25) is of fundamental importance, because it controls the mixture of fluids of all kinds, *e. g.* the solubility of gases in liquids, the miscibility and solubility of liquids with and in one another and the osmotic pressure, surface adsorption, vapour pressure and boiling points of solutions. When the law of force has been extended to the solid state, the theory will provide also a complete quantitative explanation of the solubility of solids in liquids.

§ 6. THE LATENT HEAT OF A LIQUID.

The formula in the previous section leads immediately to the calculation of the latent heat of a liquid.

If the liquid is not associated, we have from (5.26)

$$[W_s]_l^g = \frac{1}{N} (K_l v_l + P\Theta).$$

The latent heat, **L**, is the work of evaporating one gram of the liquid, which, in thermal units, is $\frac{N}{JM}$ times the work of evaporating one

$$\text{molecule, or} \quad L = \frac{1}{JM} (Kv + RT) \quad (6.1)$$

$$= \frac{R}{JM} \left(\frac{Kv}{R} + T \right)$$

In the expression (3.7) put

$$\frac{Sv}{b^{1/3} \left(v - \frac{b}{2} \right)} \equiv \chi, \quad (6.2)$$

$$\text{so that} \quad K = 4.331 \times 10^8 \chi, \quad (6.3)$$

where χ is a convenient measure of the cohesion of the liquid.

Introducing the numerical values of **J** and **R**, we have

$$\begin{aligned} L_p &= \frac{1.98}{M} \left(\frac{4.331 \times 10^8 \chi v}{R} + T \right) \\ &= \frac{1.98}{M} (5.206 \chi v + T). \end{aligned} \quad (6.4)$$

This is the value of the latent heat determined at the vapour pressure of the liquid at the temperature of the experiment. If the evaporation is carried out *in vacuo*, we have to delete the RT , obtaining

$$L_0 = \frac{1.98}{M} (5.206Xv). \quad (6.5)$$

Or, if the pressure during the experiment is any other than p , the vapour pressure, *e. g.* p_1 , we get

$$L_{p_1} = \frac{1.98}{M} \left((5.206Xv + \frac{p_1}{p} T) \right). \quad (6.6)$$

When the liquid is associated, the work of evaporating an unassociated molecule is

$$[W_{s,l}]^u = \frac{1}{N} (K_{1l}^{\frac{1}{2}} K_{12l}^{\frac{1}{2}} v_l + P\Theta) = \frac{1}{N} (A^{1/6} K_{12l} v_l + P\Theta), \quad (6.7)$$

and that of evaporating a double molecule is

$$[W_{s,l}]^d = \frac{1}{N} (K_{2l}^{\frac{1}{2}} K_{12l}^{\frac{1}{2}} v_l + P\Theta), \quad (6.8)$$

where
$$\frac{K_{2l}}{K_{1l}} = \frac{\pi \mu_2}{\sigma_2^4} \times \frac{\sigma_1^4}{\pi \mu_1} = \frac{1}{2^{1/3}}, \quad (6.9)$$

which gives
$$[W_{s,l}]^d = \frac{1}{N} \left(A^{1/6} \frac{1}{2^{1/6}} K_{12l} v_l + P\Theta \right) \\ = \frac{1}{2^{1/6}} [W_{s,l}]^u, \quad (6.10)$$

approximately.

Now, the total number of single molecules evaporated from unit area of surface per second is

$$n_1' = \frac{n_{1l} \alpha_{1l}}{2\sqrt{\pi}} e^{-\frac{s_1^2}{\alpha_1^2}}, \quad (6.11)$$

and the work done by all these molecules is

$$W_1' = \frac{W_{s,l} n_{1l} \alpha_{1l}}{2\sqrt{\pi}} e^{-\frac{s_1^2}{\alpha_1^2}}. \quad (6.12)$$

Similarly, the total number of double molecules evaporated per second is

$$n_2' = \frac{n_{2l} \alpha_{2l}}{2\sqrt{\pi}} e^{-\frac{s_2^2}{\alpha_2^2}}, \quad (6.13)$$

doing an amount of work

$$W_2' = \frac{W_{s,l} n_{2l} \alpha_{2l}}{2\sqrt{\pi}} e^{-\frac{s_2^2}{\alpha_2^2}}. \quad (6.14)$$

Now,
$$S_2^2 = \frac{2W_{s,l}}{2m}, \quad (6.15)$$

while
$$S_1^2 = \frac{2W_{s,l}}{m}. \quad (6.16)$$

So that
$$S_2^2 = \frac{1}{2^{7/6}} S_1^2 \quad \dots \quad (6.17)$$

and
$$\alpha_1^2 = \lambda_1^2 P_1^2 \quad \dots \quad (6.18)$$

$$= \lambda_1^2 \frac{4}{3} \frac{\Theta}{m}, \quad \dots \quad (6.19)$$

as
$$P_1^2 = \frac{4}{3} \frac{\Theta}{m} \quad \dots \quad (6.20)$$

by definition of temperature.

And
$$\alpha_2^2 = \lambda_2^2 P_2^2 \quad \dots \quad (6.21)$$

$$= \frac{\lambda_2^2 P_1^2}{2}, \quad \dots \quad (6.22)$$

because
$$P_2^2 = \frac{4}{3} \frac{\Theta}{2m}, \quad \dots \quad (6.23)$$

and we have seen that

$$\begin{aligned} \lambda_2^2 &= \frac{K_2^{\frac{1}{2}} K_{12}^{\frac{1}{2}} (v-b)}{P\Theta} = \frac{1}{2^{1/6}} \frac{K_1^{\frac{1}{2}} K_{12}^{\frac{1}{2}} (v-b)}{P\Theta} \\ &= \frac{1}{2^{1/6}} \lambda_1^2. \quad \dots \quad (6.24) \end{aligned}$$

So that
$$\alpha_2^2 = \frac{1}{2^{7/6}} \alpha_1^2 \quad \dots \quad (6.25)$$

and consequently

$$\frac{s_2^2}{\alpha_2^2} = \frac{s_1^2}{\alpha_1^2}. \quad \dots \quad (6.26)$$

This gives the total work done by all the single and double molecules evaporated per second from unit surface of the liquid

$$W' = W_{s_1} \frac{\alpha_{11}}{2\sqrt{\pi}} e^{-\frac{s_1^2}{\alpha_1^2}} \left(n_{11} + \frac{1}{2^{3/4}} n_{21} \right), \quad \dots \quad (6.27)$$

and the total mass evaporated under these circumstances is

$$\begin{aligned} \Sigma m &= m(n_1' + 2n_2') \\ &= m \left(\frac{n_{11} \alpha_1}{2\sqrt{\pi}} e^{-\frac{s_1^2}{\alpha_1^2}} + \frac{2n_{21} \alpha_2}{2\sqrt{\pi}} e^{-\frac{s_1^2}{\alpha_1^2}} \right) \\ &= \frac{m \alpha_{11}}{2\sqrt{\pi}} e^{-\frac{s_1^2}{\alpha_1^2}} \left(n_{11} + \frac{2}{2^{7/12}} n_{21} \right). \quad \dots \quad (6.28) \end{aligned}$$

Hence the latent heat, or the work done in evaporating one gram of liquid is, in calories, by (3.18),

$$\begin{aligned} L_{12} &= \frac{W_{s_1}}{Jm} \left(\frac{n_{11} + \frac{1}{2^{3/4}} n_{21}}{n_{11} + 2^{5/12} n_{21}} \right) \\ &= \frac{W_{s_1}}{Jm A^{7/6}} \quad \dots \quad (6.29) \end{aligned}$$

$$\begin{aligned}
 &= \frac{R}{JM} \left(\frac{A^{1/6} K_{12} v_l}{A^{7/6} R} + T \right) \\
 &= \frac{1.98}{AM} \left(\frac{K_{12} v_l}{R} + T \right) \\
 \text{very nearly,} \quad &= \frac{1.98}{AM} (5.206 \chi_{12} v_l + T). \quad \dots \dots \dots (6.30)
 \end{aligned}$$

It should be noted that χ_l corresponds to K_l with the $A^{1/3}$ in the denominator, as given in formula (3.19), while the v_l is simply the value of v in the liquid condition as formulated in (3.23). The calculated and observed values of the latent heats of the range of liquids dealt with are quoted in Table III.

In this table, the apparent volume, b , of the molecules is that calculated from the vapour pressure and given in Table IV. The association factor, A_L , is the value calculated from the latent heat. As the values of A_L agree almost exactly with those obtained from the viscosity, A_η , it will be seen that the calculated latent heats L_{12} (or L_1) agree to two significant figures with the measured values L . This confirms the fact that A is a real quantity, which is determined by the various formulæ obtained in this paper. It should be noted that when the liquid is unassociated and $A=1$, the calculated values of L_1 are identical with the observed values of L .

§ 7. THE VAPOUR PRESSURE OF AN ASSOCIATED LIQUID.

Allowing for the fact that a double molecule contains two single molecules and contributes their mass to the vapour, the total number of single molecular units evaporated per second per square centimetre of liquid surface is

$$n'_{1l} + 2n'_{2l},$$

and, in equilibrium, this is equal to the number of single molecular units from the vapour that strike the surface of the liquid in the same time. If A_g is the state of association in the vapour phase this number may be written $A_g n'_g$.

So we have

$$n'_{1l} + 2n'_{2l} = A_g n'_g. \quad \dots \dots \dots (7.1)$$

Now, as we have seen in previous papers that

$$n'_g = \frac{\alpha_g n_g}{2\sqrt{\pi}}, \quad \dots \dots \dots (7.2)$$

$$\text{we find that} \quad \frac{A_g \alpha_g n_g}{2\sqrt{\pi}} = \frac{n_{1l} \alpha_{1l}}{2\sqrt{\pi}} e^{-\frac{a_{1l}^2}{a_{1l}^2}} + \frac{2n_{2l} \alpha_{1l}}{2 \cdot 2^{7/12} \sqrt{\pi}} e^{-\frac{a_{1l}^2}{a_{1l}^2}}, \quad \dots \dots \dots (7.3)$$

where, putting

$$\begin{aligned}
 \alpha_{1l} &\equiv \lambda_{1l} P_1, \\
 \alpha_g &\equiv P_1,
 \end{aligned}$$

λ disappearing in the vapour state, we get, eliminating $\frac{\alpha_{1l}}{2\sqrt{\pi}}$,

TABLE III.

The latent heats of liquids.

Substance.	Absolute temperature T.	Molecular weight M.	Molecular volume v.	Surface tension S.	Apparent volume of molecules b.	Association factor A _L .	Latent heat.			
							Calculated.		Observed	
							L ₁ .	L ₁₂ .	L.	
Hydrogen	15° 20°	2.016	26.49 28.32	2.83 1.98	22.94 23.16	3.32 2.34	252 191	113 108.5	113 108.5	
Oxygen	70° 90°	32.00	25.9 28.0	18.3 13.2	23.33 24.05	2.51 1.88	101 77.7	55 51	55 51	
Nitrogen	70° 77.1°	28.02	33.21 34.36	10.5 8.5	28.97 29.15	2.00 1.62	78.6 65.8	49.7 47.7	49.7 47.7	
Acetone	329° 298°	58.0	77.0 73.9	19.4 22.6	66.3 65.5	1.00 1.00	125 133	— —	125 133	
Ethyl ether	307.6° 273°	74.08	106.4 100.6	14.7 19.1	91.0 89.1	1.00 1.18	84 105	— 94	84 94	
n-Hexane	341° 273°	86.02	140.1 126.9	13.4 20.5	120.5 114.5	1.06 1.35	87 117	83 91	83 91	
Benzene	358.5°	78.1	96.5	20.7	83.4	1.42	115	93	93	
Chloroform	333.2° 298°	119.5	84.5 81.0	21.8 26.2	73.2 72.6	1.16 1.55	73 85	60 63	60 63	
Carbon tetrachloride	394.4°	154	103.3	20.2	89.3	1.53	61	46	46	

$$A_g n_g = \lambda_{1l} e^{-\frac{s_1^2}{\alpha_{1l}^2}} (n_{1l} + 2^{5/12} n_{2l})$$

$$= \lambda_{1l} e^{-\frac{s_1^2}{\alpha_{1l}^2}} A_l^{5/12} n_{12l} \quad \dots \quad (7.4)$$

by (3.18).

Now
$$\frac{s_1^2}{\alpha_{1l}^2} = \frac{2W_s}{m} \div \frac{4\lambda_{1l}^2 \Theta}{3m}$$

$$= \frac{2(K_{1l}^\dagger K_{12l}^\dagger v_l + P\Theta)}{Nm} \times \frac{3}{4} \cdot \frac{\frac{2}{3}Nm}{K_{1l}^\dagger K_{12l}^\dagger (v-b)_l}$$

$$= \frac{v}{v-b} + \frac{P\Theta}{\lambda_{1l}^2 P\Theta}$$

$$= \frac{v}{v-b} + \frac{1}{\lambda_{1l}^2} \quad \dots \quad (7.5)$$

And, as
$$A_g n_g = \frac{N}{(v-b)_g} = \frac{A_g N p}{P\Theta} \quad \dots \quad (7.6)$$

and
$$n_{12l} = \frac{N}{A_l (v-b)_l} \quad \dots \quad (7.7)$$

we obtain
$$p_{\text{dynes/cm.}} = \frac{\lambda_{1l} P\Theta}{A_g A_l^{7/12} (v-b)_l} e^{-\frac{v}{v-b} - \frac{1}{\lambda_{1l}^2}} \quad \dots \quad (7.8)$$

where
$$\lambda_{1l}^2 = \frac{K_{1l}^\dagger K_{12l}^\dagger (v-b)_l}{K_l (v-b)_l} = A^{1/6} \lambda_{12l}^2.$$

So that
$$p_{\text{dynes/cm.}} = \frac{\lambda_{12l} P\Theta}{A_g A_l^\dagger (v-b)_l} e^{-\frac{v}{v-b} - \frac{1}{\lambda_{12l}^2}} \quad \dots \quad (7.9)$$

neglecting the $A^{1/6}$ multiplying the λ in the exponent because $\frac{1}{\lambda^2}$ is small compared with $\frac{v}{v-b}$.

Now, from (1.17) and (6.2) we get

$$\lambda_{12l} = 2.281 \left\{ \frac{\chi_{12l} (v-b)_l}{T} \right\}^{\frac{1}{2}} \quad \dots \quad (7.10)$$

so that, changing the temperature scale,

$$\frac{\lambda_{12l} R T}{(v-b)_l} = 1.98 \times 10^8 \left(\frac{\chi_{12l} T}{v_l - b_l} \right)^{\frac{1}{2}}$$

and
$$p_{\text{dynes/cm.}} = 1.98 \times 10^8 \left\{ \frac{\chi_{12l} T}{A(v-b)_l} \right\}^{\frac{1}{2}} e^{-\frac{v_l + 0.192 \cdot \frac{T}{\chi_{12l}}}{(v-b)_l}} \quad \dots \quad (7.11)$$

or, in mm. per sq. cm., taking 1.42 as equivalent to $\sqrt{2}$,

$$p_{\text{mm./cm.}} = \left\{ \frac{2\chi_{12l} T}{A(v-b)_l} \right\}^{\frac{1}{2}} e^{-\frac{v_l + 0.192 \cdot \frac{T}{\chi_{12l}}}{(v-b)_l}} \quad \dots \quad (7.12)$$

TABLE IV.
The vapour pressures of liquids.

Substance.	Absolute temperature T.	Molecular weight. M.	Molecular volume v .	Surface tension S.	Cohesion $K \times 10^{-8}$.	Apparent volume of the molecules b .	Association of the liquid A_L .	Vapour pressure in mm. p .	
								Calc.	Obs.
Hydrogen	15° 20°	2.016	26.49 28.32	2.83 1.98	11.3 6.7	22.94 23.16	3.32 2.34	94 689	94 690
Oxygen	70° 90°	32.00	25.9 28.0	18.3 13.2	68.4 42.8	23.33 24.05	2.51 1.88	50 756	50 760
Nitrogen	70° 77.1°	28.02	33.21 34.36	10.5 8.5	33.0 24.3	28.97 29.15	2.00 1.62	280 740	280 740
Acetone.....	298° 329°	58.00	73.9 77.0	22.6 19.4	43.7 36.6	65.5 66.3	1.00 1.00	230 760	230 760
Ethyl ether	273° 307.8°	74.08	100.6 106.4	19.1 14.7	32.1 23.8	89.1 91.0	1.18 1.00	185 760	185 760
<i>n</i> -Hexane	273° 341°	86.02	126.9 140.1	20.5 13.4	31.1 20.3	114.5 120.5	1.35 1.06	46 745	46 745
Benzene	298.1° 353.5°	78.10	89.4 96.5	27.7 20.7	40.5 33.6	81.0 83.4	— 1.42	95 760	95 760
Chloroform	298° 333.2°	119.5	81.0 84.5	26.2 21.8	44.7 38.8	72.6 73.2	1.55 1.16	203 760	203 760
Carbon tetrachloride	298° 349.4°	154.0	97.0 103.3	25.4 20.2	34.2 30.8	87.4 89.3	3.17 1.53	121 760	121 760

TABLE V.
Different values, quoted in the literature, for van der Waals' b , compared with that derived from the theory of molecular attraction.

Substance.	From theory of molecular attraction.	From viscosity.	From van der Waals' equation.	From thermal conductivity.	From maximum density.	From minimum deviation.
Hydrogen	23.7	17.5	16.6	16.2	17.5	13.3
Nitrogen	34.7	40.6	40.6	57.2	43.8	33.4

In this formula we must remember that

$$\chi_{12l} = \frac{4 \cdot 331 S_{12l} v_l}{A^{1/3} b^{1/3} (v - b)_l}, \quad \dots \dots \dots (7.13)$$

the S_{12l} being the observed surface tension and v_l the molecular volume of the liquid without any association factor.

The values of vapour pressure calculated by this formula, as quoted in Table IV., agree exactly with those observed, the value of b being taken from some other property of liquids. The effect of association in the vapour state has been omitted as being small in the cases considered. The formula may be used for calculating b for use in any one of the other formulæ, for which purpose this formula is well suited. It should be noticed that b has the same value for every property of the same liquid at the same temperature and has, like real substances, a quite small temperature coefficient of expansion. The anomaly of the large change in b , which may amount to 100 per cent. both with temperature, pressure and different properties (as in Table V.), as calculated from van der Waals' equation has now disappeared.

§ 8. THE THERMAL CONDUCTIVITY OF GASES.

A preliminary treatment of this subject will now be given. Let us suppose that heat is passed from a plane solid surface, at temperature Θ_1 , through a gas to a second plane solid surface maintained at a slightly lower temperature Θ_2 . We may suppose the gas to be arranged in a series of thin layers of equal temperature parallel to the plane of $z=0$. Let \bar{E} be the mean energy of a molecule at any plane. We have seen that \bar{E} comprises the average kinetic energy of translation of the molecules together with the internal energy of rotation of the molecule and the relative motions of its constituent atoms. And we take this internal energy as proportional to the kinetic energy, so that

$$\bar{E} = \frac{M}{3} \lambda^2 \Theta, \quad \dots \dots \dots (8.1)$$

where M is the total number of modes of motion of the molecule.

Now, the number of molecules crossing unit area of a particular plane, $z=\zeta$, per second from the direction of the surface of the solid at higher temperature, which we may call the plane $z=0$, at right-angles to the plane, is

$$v_\zeta = \frac{1}{4} n \bar{c}, \quad \dots \dots \dots (8.2)$$

where both n , and

$$\bar{c} = 4 \left(\frac{\lambda^2 \Theta}{3\pi m} \right)^{\frac{1}{2}} \cdot \frac{v}{v-b}, \quad \dots \dots \dots (8.3)$$

are functions of z , say

$$n = f_1(z), \quad \dots \dots \dots (8.4)$$

$$\bar{c} = f_2(z). \quad \dots \dots \dots (8.5)$$

Thus, we have
omitting the (z) .

$$\nu_c = \frac{1}{4} f_1 f_2, \quad (8.6)$$

The average heat energy, \bar{E} , of a molecule crossing the plane in either direction, will also depend on the position of the plane and may be written as a function of z , say

$$\bar{E} = f_3(z). \quad (8.7)$$

As in the case of viscosity, treated preliminarily in the last paper⁽¹²⁾, molecules crossing the plane, $z = \zeta$, come from an average distance $\frac{2}{3}l$, where l is the mean free path of the molecules, so that the average energy carried by the molecules crossing the plane, $z = \zeta$, corresponds to that of the planes $z = \zeta \pm \frac{2}{3}l$, according to the direction of motion

Consequently, the total heat energy carried across unit area of the plane, $z = \zeta$, per second from the direction at right-angles to the solid at higher temperature is

$$\nu_c \bar{E}_{c-\frac{2}{3}l} = \frac{1}{4} f_1(\zeta) f_2(\zeta) f_3(\zeta - \frac{2}{3}l), \quad (8.8)$$

and the total energy coming from the opposite direction is

$$\nu_c \bar{E}_{c+\frac{2}{3}l} = \frac{1}{4} f_1(\zeta) f_2(\zeta) f_3(\zeta + \frac{2}{3}l), \quad (8.9)$$

and, since

$$f_3(\zeta \pm \frac{2}{3}l) = f_3(\zeta) \pm \frac{2}{3}l f_3'(\zeta),$$

the excess of energy, Q , carried across the plane in question from the direction of the hotter solid is

$$\begin{aligned} Q &= -\frac{1}{4} f_1 f_2 f_3' \frac{4}{3} l \\ &= -\frac{1}{3} f_1 f_2 l \frac{d\bar{E}}{dz}. \quad (8.10) \end{aligned}$$

It should be noted that the $\frac{4}{3}l$ represents the sum of $\frac{2}{3}l$ taken in both directions, so that the value of l is that corresponding to the plane $z = \zeta$.

Now, the value of l is also a function of z , i. e.,

$$l = \frac{1}{\sqrt{2\pi\sigma n^2}} \quad (8.11)$$

$$= 4.35 \times 10^{-9} \frac{v}{b^{2/3}} \quad (8.12)$$

$$= f_4(z). \quad (8.13)$$

So we have

$$\begin{aligned} Q &= -f(\zeta) \frac{d\bar{E}}{dz} \\ &= -f(\zeta) \frac{M}{3} \lambda^2 \frac{d\Theta}{dz}, \quad (8.14) \end{aligned}$$

where $f(\zeta)$ is a function of the properties of the medium at the plane $z = \zeta$, and

$$\frac{d\bar{E}}{dz} = \frac{M}{3} \lambda^2 \frac{d\Theta}{dz}, \quad (8.15)$$

the small temperature change in λ being neglected.

We may write the equation (8.14) more simply as

$$Q = -k \frac{d\Theta}{dz}, \quad \dots \dots \dots (8.16)$$

where $k = \frac{M}{3} \lambda^2 f(\zeta) \dots \dots \dots (8.17)$

$$= \frac{M}{3} \cdot \frac{4}{3} \cdot \frac{N}{v} \cdot \frac{v}{v-b} \left(\frac{\lambda^2 \Theta}{3\pi m} \right)^{\frac{1}{2}} \lambda^2 \frac{v}{b^{2/3}} \times 4.35 \times 10^{-9}$$

$$= 8.94 \times 10^{26} \frac{M}{3} \left(\frac{\Theta}{Mb^{4/3}} \right)^{\frac{1}{2}} \lambda^3 \frac{v}{v-b} \dots \dots \dots (8.18)$$

This last equation shows that k is not constant, but varies with the density of the medium in each layer parallel to the surfaces of the containing solid. Experimentally, however, we measure the average effect due to the transfer of heat from one solid surface to the other. This experimental coefficient of the conduction of heat is defined by the equation

$$Q = -\kappa \frac{dT}{dz} \dots \dots \dots (8.19)$$

By comparison with (8.16) we see that

$$\kappa = \bar{k} \frac{d\Theta}{dT}$$

$$= \bar{k} \times 2.06 \times 10^{-16}, \quad \dots \dots \dots (8.20)$$

where \bar{k} is the average value of k and 2.06×10^{-16} is the constant value of $\frac{d\Theta}{dT}$, as we saw in the last paper. So, reducing to the absolute scale of T , we find

$$\kappa_{\text{ergs}} = 2.65 \times 10^3 \frac{M}{3} \left(\frac{T}{Mb^{4/3}} \right)^{\frac{1}{2}} \lambda^3 \frac{v}{v-b} \dots \dots \dots (8.21)$$

In this expression the factor $\lambda^3 \frac{v}{v-b}$ has to be averaged from its value at one solid surface to that at the other. The average is half the sum of the two end values. And as the temperatures are almost equal, the values of λ and v will be almost the same, and the average value is simply $\lambda_0^3 \frac{v_0}{v_0-b}$, where v_0 is the molecular volume of the gas as compressed or adsorbed by the much greater attraction at the solid surface, and λ_0 is the value of λ as increased by the greater attraction of the solid. The only gas considered in which the molecular volume is affected sufficiently to alter the value of $\frac{v}{v-b}$ is hydrogen. So that for other gases we need consider only λ_0 .

To get the coefficient, κ , in centigrade units we have to divide by the mechanical equivalent of heat, which gives finally

$$\kappa_{\text{centigrade}} = 6.33 \times 10^{-5} \frac{M}{3} \lambda_0^3 \frac{v_0}{v_0 - b} \frac{T^{\frac{1}{2}}}{M^{\frac{1}{2}} b^{\frac{1}{2}}} \quad (8.22)$$

It is important to realize that the values of λ_0 and v_0 in this expression are real physical quantities. When the theory of molecular attraction has been extended to the solid state, it will be possible to calculate these values exactly. We cannot do this at present. But they are substantiated by the fact that the same values fit both the equation for the thermal conductivity of gases and that for their viscosity, and this, although λ_0 and v_0 occur in very different forms in the two equations. While, also, the actual values of λ_0 are constant at different temperatures, being due, in the main, to the effect of the attraction of the solid, which is ever so much greater than that of the gas, and varies very little with temperature. The calculated values of the thermal conductivity are quoted in Table VI. with those of the viscosity at the same temperatures.

§ 9. THE VISCOSITY OF GASES.

A preliminary treatment of this subject was given by way of example in section 14 of the previous paper⁽¹²⁾. This must now be revised in the light of the facts discovered in this paper. As in § 14, *loc. cit.*, suppose the fluid to be moving with mass velocity \bar{U} in the direction OX. Consider the average momentum in this direction of molecules crossing a plane at right angles, $z = \zeta$.

The number of molecules crossing unit area of this plane per second from the direction of the stationary solid wall, which we may take as the plane $z = 0$, is

$$v_z = \frac{1}{4} n \bar{c}, \quad (9.1)$$

where both n and \bar{c} are functions of z as in the previous section.

The average mass velocity \bar{u} , in the direction OX is also a function of z , say

$$\bar{u} = f_3(z). \quad (9.2)$$

Again, as we have seen, the average distance normal to the plane $z = \zeta$, traversed by the molecules since their last encounter, before crossing the plane is $\frac{2}{3}l$, where

$$l = f_4(z). \quad (9.3)$$

Thus, the total directed momentum, in the direction OX, carried per second across unit area of the plane $z = \zeta$, from the direction of the solid, is

$$v_z m \bar{u}_{z-\frac{2}{3}l} = \frac{1}{4} m f_1(\zeta) f_2(\zeta) f_3(\zeta - \frac{2}{3}l), \quad (9.4)$$

while the total directed momentum brought from the opposite direction is

$$v_z m \bar{u}_{z+\frac{2}{3}l} = \frac{1}{4} m f_1(\zeta) f_2(\zeta) f_3(\zeta + \frac{2}{3}l). \quad (9.5)$$

Now, as

$$f_3(\zeta \pm \frac{2}{3}l) = f_3(\zeta) \pm \frac{2}{3}l f_3'(\zeta),$$

the resultant force acting on the molecules crossing the plane, which is the difference of the two expressions in (9.4) and (9.5), is

$$\begin{aligned} Z &= \frac{1}{4} m f_1 f_2 f_3' \frac{4}{3} l \\ &= \frac{1}{3} m f_1 f_2 l \frac{d\bar{u}}{dz} \\ &= \frac{1}{3} m f_1 f_2 f_4 \frac{d\bar{u}}{dz} \\ &= \frac{1}{3} m f \frac{d\bar{u}}{dz}, \quad \dots \dots \dots (2.6) \end{aligned}$$

where $f_1 \equiv f_1(\zeta)$, etc., and

$$f = f_1 f_2 f_4, \quad \dots \dots \dots (9.7)$$

being a function of the properties of the fluid in the plane $z = \zeta$.

By definition of the coefficient of viscosity, η , we have

$$Z = \eta \frac{d\bar{u}}{dz} \quad \dots \dots \dots (9.8)$$

So that

$$\eta \frac{d\bar{u}}{dz} = \frac{1}{3} m f \frac{d\bar{u}}{dz},$$

or

$$\eta_\zeta = \frac{1}{3} m f(\zeta), \quad \dots \dots \dots (9.9)$$

where $f(\zeta)$, and therefore η_ζ , depend on the state of the medium in the plane $z = \zeta$.

To evaluate the average value of η_g for a gas, we have, as before,

$$\begin{aligned} f_1 &= n = \frac{N}{v} \\ f_2 &= \bar{c} = 4 \left(\frac{\lambda^2 \Theta}{3\pi m} \right)^{\frac{1}{2}} \cdot \frac{v}{v-b} \\ f_4 &= l = 4.35 \times 10^{-9} \frac{v}{b^{2/3}}, \end{aligned}$$

giving

$$\begin{aligned} \eta_\zeta &= 4.35 \times 10^{-9} \cdot \frac{4}{3} m \frac{N}{v} \left(\frac{\lambda^2 \Theta}{3\pi m} \right)^{\frac{1}{2}} \cdot \frac{v}{v-b} \cdot \frac{v}{b^{2/3}} \\ &= 1.93 \times 10^{-5} M^{\frac{1}{2}} T \lambda \cdot \frac{v}{b^{2/3}(v-b)}. \quad \dots \dots \dots (9.10) \end{aligned}$$

This value of η_ζ is the coefficient of viscosity corresponding to the plane $z = \zeta$. But we do not measure this quantity in ordinary observations. We measure, actually, the average effect between the motionless plane $z = 0$ and the moving gas, where $z = \infty$. And the values of λ and $\frac{v}{v-b}$ have to be averaged from the planes $z = 0$ to $z = \infty$.

For a gas, there is a large increase in the value of λ at the surface of the solid, due to the much greater attraction of the molecules of the solid. The molecules of gas move much faster when they get close to

the solid, and lose this increase of velocity as they move away. Moreover, some gases, such as hydrogen, are much condensed, or adsorbed, at the surface of the solid, whereby the molecular volume of the gas is greatly decreased and the value of $\frac{v}{v-b}$ increased. The average value of $\lambda \frac{v}{v-b}$ from the state of the gas at the solid surface to that in the free gas, where λ and $\frac{v}{v-b}$ are both equal to unity is approximately

$$\lambda_0^{\frac{1}{2}} \frac{b}{v_0 - \frac{b}{2}},$$

where $\lambda_0^{\frac{1}{2}}$ is taken for convenience as equivalent to $\frac{\lambda_0}{2}$ and the values with subscript zero are the values at the solid surface. The viscosity of the gas is, therefore,

$$\eta_0 = 1.93 \times 10^{-5} \frac{M^{\frac{1}{2}} T^{\frac{1}{2}}}{b^{2/3}} \lambda_0^{\frac{1}{2}} \frac{b}{v_0 - \frac{b}{2}} \quad (9.11)$$

It is important to note that these values of λ_0 and v_0 are real physical quantities, identical with those involved in the thermal conductivity of gases. They are the necessary consequence of the greater force of attraction exerted by the solid molecules on those of the gas approaching the surface. The only alternatives are (i) a new definition of force, or (ii) to deny that the molecules of matter attract one another. It is not possible to calculate the values of λ_0 and v_0 until the law of force in the solid state has been established. But the values can be evaluated from the observations, when it is found that the values of λ_0 are constant at different temperatures, being consonant with the comparatively small thermal expansion of solids, and the values of v_0 for hydrogen are exactly inversely proportional to the quantities of hydrogen gas adsorbed by platinum sponge at the same temperatures. For example, at the temperatures 273°, 373° and 473° K., the volumes of hydrogen adsorbed by one volume of platinum sponge are respectively 3.94, 4.45 and 4.82, and the products of these with the corresponding values of v_0 are 24.7, 25 and 25. This is a further physical confirmation that v_0 is a real physical quantity. The values of λ_0 and v_0 are given in Table VI., where it will be seen that the same values fit both the thermal conductivity and the viscosity, although they enter into the respective formulæ in very different forms.

The values of the calculated and observed thermal conductivity and viscosity agree almost exactly for hydrogen, the experimental figures for which are taken from Prof. R. S. Gregory's results obtained by his new precise method. This author showed that earlier observations are subject to considerable error, which is confirmed by the greater divergence between the calculated and observed values for nitrogen and oxygen.

TABLE VI.
Thermal Conductivity and Viscosity of Gases.

Substance.	Modes of motion M.	Absolute temperature T.	Apparent volume of the molecules b.	Coefficient of thermal expansion of b between 0° and 300° K.* $\frac{db}{dT}$		λ_0	v_0	Thermal conductivity $\kappa \times 10^4$.		Viscosity $\eta \times 10^5$.	
				From κ and η .	From $v.p.$			Calc.	Obs.	Calc.	Obs.
Hydrogen ..	5	273°	23.7	0.045	0.045	1.19	52.5 50.0 45.2	4.3	4.3	8.5	8.5
		373°						5.5	5.5	10.2	10.2
		473°						6.4	6.5	12.1	12.1
Nitrogen ...	5	273°	34.7	0.025	0.025	1.25	∞	0.61	0.58	1.8	1.7
		373°						0.72	0.77	2.1	2.1
		473°						0.79	0.82	2.3	2.4
Oxygen	5	273°	31.7	0.036	0.036	1.26	∞	0.61	0.58	2.0	1.9
		373°						0.72	0.74	2.4	2.4
		473°						0.80	0.86	2.6	2.8

* Or, between 0 and 32° K. for hydrogen.

As regards the parameter λ_0 , the average kinetic energy velocities of the molecules of the gas normal to the solid surface becomes $\frac{1}{2}m\lambda_0^2C^2$, when the molecules collide with those of the solid. In rebounding after collision, the velocity of the molecules of the gas falls off very quickly, through the distance of only a few molecular diameters, to the value $\frac{1}{2}mC^2$ in the body of the gas. During this extremely short distance, the value of C and, therefore, of the temperature, $\frac{1}{2}mC^2$, has changed very little. In thermal conductivity experiments, the temperature changes at a constant rate from one solid surface to the other. The rapid fall in energy corresponds to a change that has been called "temperature drop." It is not a change in temperature, however, but one of kinetic energy, due to the falling off of the attraction of the molecules of the solid.

An exactly similar rapid diminution of average kinetic energy takes place in the phenomenon of viscosity. Only, in this, the effect is represented by $\lambda_0^{\frac{1}{2}}$ instead of λ_0^3 . This smaller effect has been diagnosed as "accommodation." But Table VI. shows that the same value of λ_0 fits both phenomena, thermal conductivity and viscosity at all temperatures. Moreover, it must be remembered that, in this theory of molecular attraction, all collisions have been assumed to take place between perfectly elastic molecules, not merely between a *molecule* of gas, or liquid, and a solid *surface*. Since this supposition agrees with the observed facts, it is unnecessary to introduce an additional elastic factor, which has to be chosen *ad hoc* for every substance, temperature and property.

The values of b and its temperature coefficients are the same for both properties of gases under discussion. They appear to reach a maximum at 32° K. for hydrogen and 300° K. for oxygen and nitrogen. The temperature coefficients agree exactly with those at very much lower temperatures, for all properties discussed, of the same substances in the liquid state. Thus we see again that the parameter b represents a definite physical property of matter, which is always the same at the same temperature for the same substance. The great difference between these definite values of b and those deduced from previous treatments, by other hands, of various properties of gases has been shown in Table V. In connexion with this table it will be recollected that van der Waals' b varies by as much as 100 per cent. with change of pressure, or temperature.

§ 10. THE VISCOSITY OF LIQUIDS.

Turning now to the viscosity of liquids, we realize that the increase in the value of λ at the solid surface in a liquid is not to be compared with that in a gas, because the mutual attraction of the molecules of the liquid is so much more equal to that of these for the solid molecules. To simplify the equations, it is permissible to neglect this increase in λ as a separate quantity and to suppose the whole effect of the greater attraction of the solid to be concentrated in an increase of density of

the liquid at the solid surface. This latter change has, in fact, more effect in a liquid, because the value of v begins to approach that of b . As for gases, we cannot calculate the effect at present from the law of attraction in the solid state. But we can find, from the observed figures for the viscosity of a number of liquids, a most probable form of the factor $\frac{v}{(v-b)^{\frac{1}{2}}}$, deduced from $f_1 f_2 f_4$, which fits all exactly. This form,

i. e. $\frac{v+3b}{(v-b)^{\frac{1}{2}}}$, contains only the molecular volume of the liquid itself and its apparent volume of the molecules b , so that the value of the factor in each case can be obtained from the published figures and the value of b that fits another property of a liquid at the same temperature.

So for the viscosity of a liquid we have

$$\begin{aligned} f_1 &= n = \frac{N}{v_l}, \\ f_2 &= \bar{c} = 2\sqrt{2} \left\{ \frac{K v_l^2}{\pi M (v-b)_l} \right\}^{\frac{1}{2}}, \\ f_4 &= l = 4.35 \times 10^{-9} \frac{v_l}{b^{2/3}}. \end{aligned}$$

Consequently

$$\eta_l = 4.8 \times 10^{-5} \frac{v_l + 3b}{2b^{2/3}} \left\{ \frac{X_l M}{v_l - b} \right\}^{\frac{1}{2}}, \quad \dots \quad (10.1)$$

where the v_l is the molecular volume in the liquid state.

We must observe, however, that the value of l corresponds in reality to that of Tait, rather than that of Maxwell, because we are dealing with the average distance travelled by a molecule since its last collision and a given instant. This introduces a factor 0.958, which gives finally

$$\eta_l = 2.3 \times 10^{-5} \frac{v_l + 3b}{b^{2/3}} \left\{ \frac{X_l M}{v_l - b} \right\}^{\frac{1}{2}}, \quad \dots \quad (10.2)$$

This is the equation of the viscosity of an unassociated liquid. But few liquids are entirely unassociated and association necessarily affects their viscosity.

To determine the effect of association in viscosity, we have only to consider the motions of the single and double molecules separately.

The number of collisions of one of the single molecules per second with all the n_1 single molecules is

$$B_{11} = \sqrt{2} \sqrt{\pi} n_1 \sigma_1^2 \alpha_1, \quad \dots \quad (10.3)$$

while the number of collisions of a single molecule with all the n_2 double molecules is

$$B_{12} = \sqrt{\pi} n_2 \sigma_{12}^2 (\alpha_1^2 + \alpha_2^2)^{\frac{1}{2}}, \quad \dots \quad (10.4)$$

and we have seen that

$$\alpha_2 = \frac{1}{2^{7/12}} \alpha_1,$$

$$\sigma_2 = 2^{1/3} \sigma_1,$$

and

$$\sigma_{12} = 2^{1/6} \sigma_1.$$

Consequently

$$B_{12} = \sqrt{\pi n_2 \sigma_1^2 \alpha_1} \left(1 + \frac{1}{2^{7/6}}\right)^{1/2} 2^{1/3} \\ = 1.514 \sqrt{\pi n_2 \sigma_1^2 \alpha_1}. \quad (10.5)$$

Similarly

$$B_{21} = 1.514 \sqrt{\pi n_1 \sigma_1^2 \alpha_1}, \quad (10.6)$$

and

$$B_{22} = \sqrt{2} \sqrt{\pi n_2 \sigma_2^2 \alpha_2} \\ = \sqrt{2} \sqrt{\pi n_2 \sigma_1^2 \alpha_1} \frac{2^{2/3}}{2^{7/6}}, \\ = \sqrt{\pi n_2 \sigma_1^2 \alpha_1}. \quad (10.8)$$

The mean path, l_1 , of the single molecules is the mean speed, \bar{c}_1 , divided by twice the total number of collisions per second of the single molecules. So that

$$l_1 = \frac{\bar{c}_1}{2(B_{11} + B_{12})}, \quad (10.9)$$

where

$$\frac{\bar{c}_1}{\alpha_1} = \frac{\bar{c}_2}{\alpha_2} = \frac{2}{\sqrt{\pi}}, \quad (10.10)$$

by Maxwell's theory.

$$\therefore l_1 = \frac{1}{\pi \sigma_1^2 (1.414 n_1 + 1.515 n_2)} \quad (10.11)$$

and

$$l_2 = \frac{1}{\pi \sigma_1^2 (1.515 n_1 + n_2)}. \quad (10.12)$$

The n_1 single molecules will contribute to the force exerted per unit area of the plane

$$Z_1 = \frac{1}{4} m_1 \bar{c}_1 n_1 \frac{4}{3} l_1 \frac{d\bar{u}}{dz} \quad (10.13)$$

$$= \frac{1}{3} n_1 m_1 \bar{c}_1 \frac{du}{dz} \cdot \frac{1}{\sqrt{2} \pi \sigma_1^2 (n_1 + 1.07 n_2)}, \quad (10.14)$$

while the n_2 double molecules will exert a force

$$Z_2 = \frac{1}{4} 2 m_1 \bar{c}_2 n_2 \frac{4}{3} l_2 \frac{d\bar{u}}{dz} \quad (10.15)$$

$$= \frac{2}{3} n_2 m_1 \bar{c}_2 \frac{du}{dz} \cdot \frac{1}{\sqrt{2} \pi \sigma_1^2 (1.07 n_1 + 0.692 n_2)}, \quad (10.16)$$

where

$$\bar{c}_2 = \frac{1}{2^{7/12}} \bar{c}_1.$$

So we obtain the total force

$$Z = Z_1 + Z_2 = \frac{1}{3} m_1 c_1 \frac{du}{dz} \cdot \frac{1}{\sqrt{2} \pi \sigma_1^2} \left(\frac{n_1}{n_1 + n_2} + \frac{\frac{2}{2^{7/12}} n_2}{n_1 + \frac{n_2}{2^{1/2}}} \right), \quad (10.17)$$

neglecting the factor 1.07,

In the denominator of the second fraction in parentheses we have

$$n_1 + \frac{n_2}{2^{\frac{1}{2}}} = \frac{n_1 + n_2}{A^{\frac{1}{2}}}, \quad \dots \quad (10.18)$$

that we find

$$Z = \frac{1}{3} m_1 \bar{c}_1 \frac{du}{dz} \cdot \frac{1}{\sqrt{2\pi\sigma_1^2}} \left(\frac{n_1}{n_2 + n_2} + \frac{2^{5/12} n_2}{n_1 + n_2} \right) \quad \dots \quad (10.19)$$

approximately, because

$$A^{\frac{1}{2}} = 1 + \frac{1}{2} h,$$

if h is not too large a fraction, and

$$\begin{aligned} \frac{2}{2^{7/12}} A^{\frac{1}{2}} &= \frac{2}{2^{7/12}} + \frac{1}{2^{7/12}} h \\ &= 2^{5/12}. \end{aligned}$$

This gives
$$Z = \frac{1}{3} m_1 \bar{c}_1 \frac{d\bar{u}}{dz} \cdot \frac{A^{5/12}}{\sqrt{2\pi\sigma_1^2}} \quad \dots \quad (10.20)$$

Now
$$\bar{c}_1 = \frac{2}{\sqrt{\pi}} \alpha_1 = \frac{2}{\sqrt{\pi}} \left(\lambda_1^2 \cdot \frac{4}{3} \frac{\Theta}{m_1} \right)^{\frac{1}{2}}, \quad (10.21)$$

where λ_1 involves $K_1^{\frac{1}{2}} K_2^{\frac{1}{2}}$ and introduces a factor of $\frac{1}{A^{1/12}}$, if we take K_1

in the final formula for the viscosity. Consequently the effect of association on viscosity is to introduce a factor $A^{1/3}$, which, however, increases slightly with increasing A . For values up to about $A=1.6$, the factor $A^{0.36}$ gives accurate results. Above that, the index of A becomes larger and the calculated values of A derived from the factor $A^{0.36}$ are greater than the real values derived from the latent heat formula. The viscosity of an associated liquid is, therefore,

$$\eta_l = 2.3 \times 10^{-5} A^{1/3} \frac{v_l + 3b}{b^{2/3}} \left\{ \frac{\chi_l M}{v_l - b} \right\}^{\frac{1}{2}} \quad \dots \quad (10.22)$$

Except for this unavoidable divergence with high values of A , the values of this quantity calculated from the viscosity formula agree exactly with those obtained from the latent heat. And the values of the viscosity itself are accurate to two significant figures, as is shown in Table VII. It should be mentioned that this formula for the viscosity does not apply to a liquid like mercury, which has a cohesion equal to, or greater than, that of the glass measuring apparatus. The coefficient $\frac{v+3b}{(v-b)^{\frac{1}{2}}}$ will no longer be greater than $\frac{v}{(v-b)^{\frac{1}{2}}}$ but rather less, because the molecular volume at the interface will not be contracted by a greater attraction of the solid molecules. It would be easy to find the value of the coefficient, which would give accurate values of the viscosity of mercury at all temperatures. But this would be merely one case of an exceptionally heavy liquid and would have no particular significance from the point of view of the general theory.

TABLE VII.
The Viscosities of Liquids.

Substance.	Absolute temperature T.	Association factor A_{η} .	Viscosity η .	
			Calculated.	Observed.
Acetone.....	329.0°	1.00	0.0 ₂ 26	0.0 ₂ 24
	298.0°	1.00	0.0 ₂ 32	0.0 ₂ 32
Ether.....	307.6°	1.00	0.0 ₂ 23	0.0 ₂ 23
	273.0°	1.18	0.0 ₂ 29	0.0 ₂ 29
<i>n</i> -Hexane	341.0°	1.05	0.0 ₂ 22	0.0 ₂ 22
	273.0°	1.35	0.0 ₂ 40	0.0 ₂ 40
Benzene	353.5°	1.44	0.0 ₂ 32	0.0 ₂ 32
Chloroform	333.2°	1.10	0.0 ₂ 41	0.0 ₂ 41
	298.0°	1.56	0.0 ₂ 59	0.0 ₂ 59
Carbon tetrachloride	339.4°	1.80	0.0 ₂ 50	0.0 ₂ 50

§ 11. CONCLUSION.

In this series of papers, a general theory of the properties of fluids has been developed, in which molecular attraction plays a rôle, identical with that of gravitation in astronomy. Every property of fluids considered has been calculated directly, without the introduction of any arbitrary constants, by means of Newton's dynamics from Maxwell's kinetic theory and Edser's inverse eighth power law of molecular attraction. The exact agreement to two significant figures of the calculated and observed values is as near as can be expected, and provides conclusive proof that Newton's laws apply to the motions of molecules. The equations developed are not in any way a contradiction of the accepted dynamical theory of gases, for which the value of λ in the new equations reduces to unity and the equations themselves to the accepted equations for gases. The new theory is at once a confirmation of the dynamical theory of gases and an extension of it to the liquid state.

It is clear that the method can be developed further to include the solid state by necessary modification of the law of force. The same method must give also an exponential law of radiation. As yet, however, much remains to be done in the theory of fluids. Some points discussed in this paper need further examination on the lines indicated. The difficult problem of trebly associated water needs to be resolved. The physical nature of solution, as the evaporation of solute into solvent, requires further study. This will give a quantitative explanation of the process of solution and enable the calculation of the actual solubility in different solvents, of gases, liquids and solids. Osmosis, depression of

the freezing point, elevation of the boiling point, electrolysis, viscosity and all the properties of solutions will certainly appear, as shown in previous papers, to be the result of attraction, volume, mass and motion only. The author would not wish to restrict so much work to his own hand; so wide a subject needs many workers, whose collaboration is invited.

The properties of colloids have been studied already in the light of the general principles now worked out in detail for fluids. A theory of gels was published in the 'Biochemical Journal' ⁽⁴⁾, and previous papers. In the volumes quoted it was proved, by actual crystallization of gelatin and in other ways, that the process of gelation is the same as that of crystallization, with the only difference that definite conditions determine the production of smaller particles. A note was added to that paper to the effect that specially pure isoelectric gelatin did not yield crystals so readily. Shortly afterwards, spherites of isoelectric gelatin were obtained of smaller size, as is now here recorded. Agar agar proved more difficult to crystallize, because of its smaller solubility. The method of recrystallization was adopted. A microscopically clear jelly of pure dialysed agar agar, 1.2 per cent., was sealed in boiling tubes on 13th January, 1921, and kindly autoclaved by Prof. Paine in the Imperial College of Science and Technology. The tubes were opened on 29th June, 1940, after 19½ years. The agar jelly had become opalescent and, even with a $\frac{2}{3}$ -inch objective, the jelly could be seen to be completely full of grains. With $\frac{1}{8}$ -inch and $\frac{1}{12}$ -inch objectives, the particles were spherical and just under 1μ for the most part. A number reached 1μ or 1.5μ about. These larger grains gave distinct shadow curves with crossed Nicols. The whole appearance resembled exactly that of the precipitate from very slightly supersaturated solutions of gelatin maintained at a constant temperature, as described in the papers mentioned, although the agar agar particles were rather larger. During this long period the invisible agar agar particles had gradually grown by recrystallization into spherites of microscopic size. The experiments on the crystalline nature of gelatin gels have been confirmed by other hands by X-ray analysis. The way is now open for quantitative development of the theory of gels.

References to Previous Papers on this Subject.

- (1-4) "On the Theory of Gels" [I] (-IV), *Biochem. J.* xii. pp. 351-381 (1918); xiv. pp. 91-93 (1920); xv. pp. 553-562 (1921); xvii. pp. 230-239 (1923).
- (5) "The Reversible Sol-gel Transformation." Faraday Society, Symposium, pp. 44-50 (1920).
- (6) "The Sol-gel Transformation and the Properties of Jellies." Alexander, J., 'Colloid Chemistry,' i. pp. 753-766 (1926).
- (7-8) "On the Molecular Theory of Solution" [I] (-II), *Phil. Mag.* xxxviii. pp. 696-705 (1919).
- (9-11) "Contribution to the Kinetic Theory of Vaporization" [I] (-III), xlviii. pp. 936-946 (1924); l. pp. 1147-1158 (1925); x. pp. 160-185 (1930).
- (12) "The Kinetic Theory of Fluids." *Proc. Phys. Soc.* l. pp. 30-58 (1938).

LIII. *The Problem of the Keyboard Instrument.*

By LL. S. LLOYD*.

[Received March 15, 1943.]

IN some books on Sound written towards the end of the nineteenth century it is stated that a cycle of fixed tones corresponding to vibrations with frequencies in the proportions

$$1, \frac{9}{8}, \frac{5}{4}, \frac{4}{3}, \frac{3}{2}, \frac{5}{3}, \frac{15}{8}, 2$$

is a musical scale that is satisfactory till it becomes necessary to modulate. The inadequacy of this cycle as a musical scale is independent of modulation. It is due to inherent defects in the structure of the cycle. Some nineteenth-century theorists even called the cycle "just intonation": a misuse of an established technical term that has caused much mischief, as I have explained elsewhere†. That is why some modern writers, by way of protest, are careful to call it "just temperament." The most obvious inadequacy of the cycle is easily discovered. Call the notes of the series *c, d, e, f, g, a, b, c'*. The frequency ratio of *d a* is

$$\frac{5}{3} \div \frac{9}{8} = \frac{40}{27} = \frac{3}{2} \div \frac{81}{80}.$$

Similarly, the frequency ratio of *d f* is

$$\frac{4}{3} \div \frac{9}{8} = \frac{32}{27} = \frac{6}{5} \div \frac{81}{80}.$$

In a musical scale whose intonation is perfect (just), these intervals should be a perfect fifth ($3/2$) and a minor third ($6/5$). Actually each is too small by a comma ($81/80$). But if *d* be moved down a comma, till its frequency ratio with *c* is $10/9$, the triad *d f a* is true. In a musical scale *d* becomes a mutable note, as Stanford explains at the beginning of his 'Musical Composition.'

The misconception with which we began appears to have first arisen in the latter part of the nineteenth century. Older writers knew better. Thus, in a paper "On the division of the Monochord, or Section of the Musical Canon" in the *Phil. Trans.* 1698, Dr. John Wallis divided the "ditone, or greater third," whose ratio he took as $5 : 4$, into the greater tone, $9 : 8$, and the lesser tone, $10 : 9$. "But whether *fa*, *sol* [*c, d*] shall be made the lesser, as 10 to 9, and *sol*, *la* [*d, e*] the greater, as 9 to 8; or

* Communicated by Sir Edward Appleton, F.R.S.

† "Just Intonation Misconceived," 'Music and Letters,' July 1943.

this the lesser, as 10 to 9, and that the greater, as 9 to 8; or sometimes this and sometimes that, as there is occasion . . . is somewhat indifferently." It may be remarked that, in the *gam* beginning with *G* on the bottom line of the bass stave as *ut*, *c* was named *fa*, *d* *sol* (or *d sol re*) and *e la* (or *e la mi*)*. Here, then, we have a description, 250 years old, of the mutable notes which Stanford explained for the instruction of his students; and, as he also explained, in the relative minor key of *A*, *b* and *g* are mutable notes. Nor is this all the evidence of the flexible nature of the scale-system of music, as we shall see.

An even more apposite passage may be quoted from Dr. Robert Smith's † 'Harmonics' (1st ed. 1748, 2nd ed., from which I quote, 1758), page 32:

"Tune upwards from *C* the two perfect V^{th} *CG*, *Gd*, and the perfect $XVII^{th}$, or 2 $VIII+III$, *Ce'*, then downwards the V^{th} *e'a*, and the intermediate fifth *ad* will be too little by a comma, as including the imperfect minor third *df*. And by tuning an eighth below *a* we have the imperfect fourth *Ad* too large by a comma.

"The disagreeable effect of this fifth *da* and the fourth *dA* in every octave, and of their compounds with the $VIII^{th}$, and also of the third *df* and the sixth *fd'* in every octave and of their compounds with the $VIII^{th}$, and of many more such imperfect concords, when the usual flat and sharp sounds are added to complete the scale, has obliged practical musicians, long ago, to distribute that comma, wanting in the fifth *da*, equally among all the four V^{th} s, *CG*, *Gd*, *da*, *ae'*, contained in the $XVII^{th}$ *Ce'*. And this interval *Ce'* may be increased or decreased a little before it be divided into 4 equal V^{th} s. In any case such distribution is therefore called the Participation or Temperament of the system, and when rightly adjusted is undoubtedly the finest improvement in harmonics."

This presents the problem of tuning keyboard instruments as being concerned solely with the defects in the cycle with which we began.

The system of tuning described by Dr. Smith is mean-tone temperament, the "vulgar" tuning in general use in his day. It was originally devised in the sixteenth century, for modal music, and the "scale" of modal polyphony was surprisingly flexible‡. It ultimately gave place to equal temperament, as a method of tuning keyboard instruments to meet the demands of classical composers for a wide range of key and more varied chromatic usages, of which free modulation is one aspect. But the problem of the keyboard instrument is much older: observe Dr. Smith's "long ago." It is: to devise that system of tuning an instrument of fixed intonation which gives the best approximation to the flexible

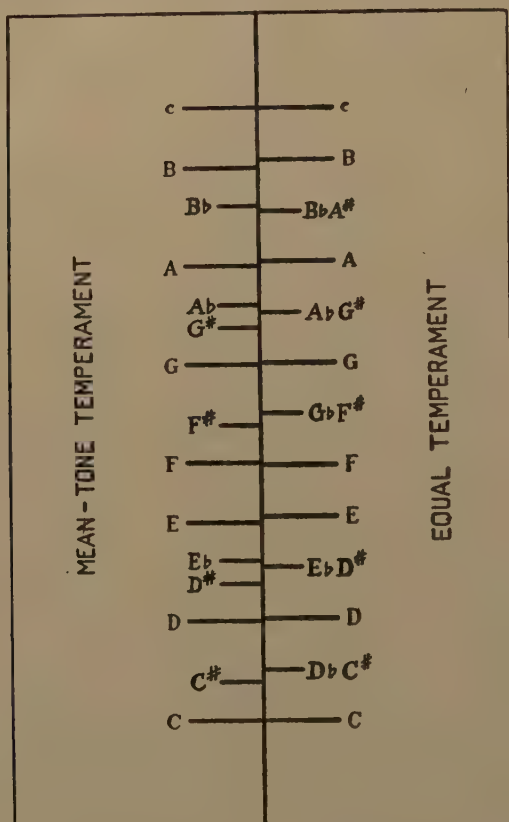
* See Morley's 'Plaine and Easie Introduction to Practicall Musicke' (1597).

† Dr. Robert Smith, F.R.S., Master of Trinity College, Cambridge, was the founder of the Smith's prizes.

‡ See e. g. "The Musical Scale," 'The Musical Quarterly,' April, 1942, by the present writer.

intonation of the music of any given period. But it must always result in what, for delicate ears, is a makeshift.

Mean-tone temperament is so important in considering the attempts at solution of this problem that it may be described rather more fully. It is contrasted with equal temperament in the accompanying figure. Its essential features are three. Like all practical temperaments, it begins by eliminating the difference between the major tone ($9/8$) and the minor tone ($10/9$). It flattened all the fifths, and therefore increased all the



fourths, by $\frac{1}{4}$ comma, and it made all the major thirds perfect. The true major third ($5/4$) is the sum of a major tone and a minor tone. Mean-tone temperament therefore made all whole tones the mean of these two intervals, $\frac{1}{2}$ comma smaller than the major tone, $\frac{1}{2}$ comma larger than the minor tone. The black keys were tuned to give the chromatically altered notes used in modal music: C \sharp , E \flat , F \sharp , G \sharp , B \flat . C \sharp was tuned to be a perfect major third above A, F \sharp a perfect major third above D, while E \flat was a perfect major third below G. The tuning of the black notes being determined in this way by that of the white ones, it is easy to see why this tuning, evolved in the sixteenth century, was afterwards found to be suitable for only a few keys. G \sharp , a major third above E, was nearly two

commas too low to form the perfect major third $A\flat C$, while $E\flat$ was nearly two commas too high for use as $D\sharp$. This produced "wolves" in the keys of $E\flat$ major and E major, and in all the major keys more remote from C , while the minor keys of C and E were faulty. That is why on some organs the black notes for $E\flat$ and $G\sharp$ were divided as indicated in the figure, and two additional ranks of pipes inserted in each octave to correspond with the notes so added for $D\sharp$ and $A\flat$.

It is evident from the figure that $A\flat$ lies above $G\sharp$ in mean-tone temperament. The same is true of just temperament, and the interval $G\sharp A\flat$, the "diesis" of this temperament, is the same in both temperaments. But in just temperament, using the black notes for the chromatic notes of the key of C , the interval $F\sharp G\flat$ is a comma less. This brings us to another difficulty, which exhibits a further inadequacy of just temperament as a substitute for the perfect (just) intonation of stringed instruments. Moritz Hauptmann, a distinguished violinist whose theoretical work was referred to by Helmholtz with much respect, had this to say*:

"It is a matter of easy calculation that, in respect of C , $C\sharp$ is lower than $D\flat$; but if I use $C\sharp$ as a leading note, and $D\flat$ as a minor ninth, my first note is far ahead of the other Had I no natural consciousness of the pure tone how could I recognise the sharpening of it as the leading note—its flattening as a minor ninth?"

In other words, $C\sharp$ and $D\flat$ have now changed places. Hauptmann is here writing of decorating notes which do not form part of the prevailing harmony, when their intonation will be determined by the player's melodic feeling.

Hauptmann's criticism of the piano, as an instrument to accompany the violin, was not of its *faulty* intonation, due to the mistuning required for temperament, but of its *rigid* intonation. He complained that the intonation of the violin had to give way to that of the piano, a complaint which must have referred, primarily, to consonances, and one that would come naturally from a violinist accustomed to the sensitive response of a string quartet in matters of intonation.

Even this is not the end of the story; for as Helmholtz explained, scientifically, concords have far more "definition," to the ear, than discords. A small shift of intonation of one note of a concord readily disappears in a succeeding discord. Helmholtz pointed out that these "enharmonic changes," as the musician calls them, are best made in discords such as the chord of the diminished seventh, whose unstable intonation made it so useful in an enharmonic modulation. But Dr. Robert Smith was well aware of this. In explanation of an improved temperament which he proposed, he wrote that he had only paid regard to concords, "because the ear is generally less critical in the discords than in the concords."

It was unfortunate that, obsessed by the problems of euphony pre-

* 'Letters of a Leipzig Cantor' (Novello), vol. i. p. 150.

sented by keyboard instruments, nineteenth century theorists were unaware how naturally all these problems had been solved, for unaccompanied voices, by Palestrina in his compositions. For in their day the contrapuntal technique of the sixteenth century had been lost to sight because of the harmonic fog which then arose from thinking of chords as things existing in themselves*. The resulting misconceptions were more fully discussed in the paper already referred to†. Perhaps the fundamental mistake of nineteenth-century theorists was their assumption that in musical performance a note, once sounded, must be maintained with a fixed intonation so long as it lasts. That is not true, as is evident from the simplest examples in Palestrina's writing. Blinded by this mistake, they assumed that in certain passages faulty *concord*s would be produced, or that there would be constant alterations of pitch by a comma or more. But Huygens knew better, as Dr. Robert Smith reminds us in his 'Harmonics,' p. 228 :

"Mr. Huygens observed long ago, that no voice or perfect instrument can always proceed by perfect intervals, without erring from the pitch at first assumed. But as this would offend the ear of the musician, he naturally avoids it by his memory of the pitch, and by tempering the intervals of the intermediate sounds, so as to return to it again.

"This is also confirmed by what we are told of a monk, who found, by subtracting all the ascents of the voice in a certain chant from all its descents, that the latter exceeded the former by two commas : so that if the ascents and descents were constantly made by perfect intervals, and the chant were repeated but four or five times, the final sound, which in that chant should be the same as the initial, would fall about a whole tone below it. But finding that the voices in his choir did not vary from the pitch assumed, he concluded that the musical ratios, whereby he measured those successive ascents and descents, were erroneous. But if he had known Mr. Huygens' remark, it would have solved his difficulty."

And we shall not read into this any meaning not intended by Dr. Smith if we turn to cor. 3 of proposition XXII. which precedes it :

"*Cæteris paribus* the same piece of music well performed upon perfect instruments, is more agreeable than it would be if it were as well performed upon imperfect ones, as an organ, etc."

And to proposition XXII. itself, where we find :

"For the passing from one sound to the next, whether by a perfect or an imperfect interval, being nearly instantaneous, cannot much offend the musician. But the succeeding consonance is long enough held out to give him pleasure or pain according as he makes it perfect or imperfect. Q.E.D."

* Cf. the article "Harmony" by Sir Donald Tovey in Enc. Brit. 14th ed.

† "Just Intonation Misconceived," 'Music and Letters,' July 1943.

Perhaps the quotations above are sufficient to show that, by bringing his conclusions to the test of accurate observation, Dr. Robert Smith, though lacking all the knowledge of the structure and properties of the ear which Helmholtz had at his disposal, was able to anticipate much that his successor had to say. Careful study of his 'Harmonics' leaves the conviction that Dr. Robert Smith's book is an original contribution to musical acoustics that deserves to rank with 'Tonempfindungen' as a landmark in the history of the subject. Perhaps because it is stiff reading, perhaps because he communicated no paper on his work to the Phil. Trans.*, this remarkable book has not received the recognition it deserves. Its immediate and practical purpose was to improve the temperaments of keyboard instruments. But its most notable achievement is that, proceeding from purely mathematical premises, it discovered, accurately, those rates of beating produced in mistuned intervals which were determined, with reference to the properties of the ear, more than a century later, by Helmholtz. From the criticism Dr. Smith offered of previous theories about beats, notably that of Sauveur, it is evident that he was breaking virgin ground. His book is a triumph of the inductive method; for starting with a postulate, of which he writes "I found it necessary to premise a Theory of Imperfect Consonances," by purely mathematical means he reaches conclusions which he then proceeds to bring, successfully, to the test of experimental verification. He seems to come within an ace of anticipating Helmholtz's theory of dissonance, yet because he knew nothing of the ear his premise is really "speculative," so that there was still a long way to travel. In his 'Plaine and Easie Introduction to Practicall Musicke,' 150 years earlier, Thomas Morley described two kinds of music: and of the first he wrote:

"*Speculative* [musicke] is that kinde of musicke which by Mathematical helps, seeketh out the causes, properties, and natures of soundes by themselves, and compared with others proceeding no further, but content with the onlie contemplation of the Art."

It seems to me that in Dr. Robert Smith's 'Harmonics' these "Mathematical helps" reached their high water mark. Until the knowledge that Helmholtz laid under contribution became available there was little or no room for any further advance along the road used by these "helpes." Even d'Alembert, who reduced Rameau's speculations to an orderly system, did not make use of beats—as Helmholtz pointed out, observing that they were "the real source of the distinction between consonance and dissonance."

Here, for example, is an illustration of the experimental use made of them by Dr. Robert Smith. *Scholium* 2 of his Proposition XI. has the heading *To show that the theory of beats agrees with experiments*, and contains the following paragraph, p. 93:

* Dr. Robert Smith is not mentioned in 'Tonempfindungen.'

"Hence any two imperfect consonances [*i. e.*, mistuned consonances] which compose a perfect 8th, will be equally quick, if the minor consonance be below the major; but if the minor be above the major, it will beat twice as quick as the major."

This, though not generally known save in the special case of the fifth and the fourth (regarded as major and minor intervals), is perfectly true. Suppose a perfect major third to be tuned up from a given note, and a perfect minor sixth to be tuned down from it, making a true octave. Now mistune the "given note" by one vibration a second. As Helmholtz showed, the rate of beating between the harmonics of the two notes forming the major third (ratio 5 : 4) will be 5 (the larger number of the ratio when the lower note is mistuned one vibration a second). The rate of beating between the harmonics of the two notes forming the minor sixth (ratio 8 : 5) will also be 5 (the smaller number of the ratio when the upper note is mistuned one vibration a second). But if the intervals are now interchanged, the minor sixth will give 8 beats and the major third 4 beats a second. And for the mistuning of the "given note" by any small quantity x , these rates will become $5x$ in the first case, and $8x$ and $4x$ in the second.

In computing, by geometrical means, the rate of beating of a tempered fifth, Dr. Robert Smith in effect uses the third harmonic of the lower note and the second harmonic of the higher note, which is what Helmholtz did. But as harmonic overtones were not used in his mathematical argument, he was under no necessity to express his observation in terms of them; though as he was familiar with Sauveur's work, he was presumably aware of the existence of harmonic overtones in a note. It is, however, unlikely that he appreciated their musical significance, for he never mentions them. The omission may or may not be worth exploring, for he clearly took his title from Ptolemy.

The "theories" of the nineteenth century about music and sound were mostly armchair theories. They were fabricated by thinking about them, but they were seldom brought to the test of either laboratory experiment or musical history as known to scholars to-day. It is therefore to be regretted that obsession with their own theories should have led nineteenth-century theorists to ignore so much good work of earlier date. Nor does the mischief they did end there. This obsession with "theories" that we rightly ignore to-day led the English translator of Helmholtz's '*Tonempfindungen*' to read meanings into his translation of Chapter XVI. which have misled many English readers. Those who wish to judge of the validity of this criticism will find the evidence in the paper already referred to*. Thirty-three passages, most of them doubtful, and an example of the translator's inadequate knowledge of musical technique, are quoted in that paper by kind permission of Messrs. Longmans, Green and Co., the publishers of '*Sensations of Tone*,' for which the writer records his cordial thanks.

* "*Just Intonation Misconceived*," '*Music and Letters*,' July 1943,

The search for an ideal temperament will always be as fruitless as the search for the philosopher's stone. But it is of more importance that this is, in fact, of very little moment : for two reasons. First, instrumental music moves with greater rapidity, in general, than the vocal music of the sixteenth century. As Helmholtz observed, and, as Dr. Smith knew, as we have seen, this makes intonation that is imperfect, as measured by physical vibrations, less offensive to the ear. Second, the defects of temperaments are not so obvious on some keyboard instruments as on others. In brief, on the piano they are of little or no importance except for those with the good ear of the really skilled violinist. On the organ in a cathedral, with its long period of reverberation, they are in a large measure disguised by the acoustics of the building. But they are fatally objectionable in the sustained and incisive tones of the harmonium, though, fortunately, the problem is then susceptible of an obvious solution.

In conclusion the writer wishes to express his indebtedness to the Oxford University Press for permitting the use of the figure herein, which is based on the 'Musical Slide Rule' that accompanies his booklet with that title.

Department of Scientific and
Industrial Research.

LIV. *Three Types of Energy of Viscosity.*

By A. E. DUNSTAN, D.Sc. and A. H. NISSAN, Ph.D.*

[Received December 24, 1942.]

FOR comparatively narrow ranges of the temperature the viscosity-temperature relationship of unassociated liquids is given by

$$\log_e \mu = A' + \frac{B}{T}, \quad (1)$$

where

μ = viscosity in absolute units,

T = temperature on the absolute scale,

A' & B = constants.

The "constant" B has been shown to be a function of temperature⁽¹⁾, but when the range is small it may be assumed to be invariant. This function is usually subdivided into a ratio of two terms,

$$B = \frac{Q}{R},$$

where

Q = viscosity energy of activation,

R = gas constant.

* Communicated by the Authors.

Thus
$$\log \mu = A + \frac{Q}{2.3RT} \quad \dots \dots \dots (2)$$

An equation has been found ⁽²⁾ giving the viscosity temperature relationship of the normal paraffins, from C₄ to C₁₈ inclusive, covering the entire temperature range over which the substances exist as liquids under atmospheric pressure. This equation is of the same form for all the normal paraffins, excepting the first three members, and is

$$\log \mu = 6.283 - 18.572 \left(\frac{T}{T_B} \right) + 17.505 \left(\frac{T}{T_B} \right)^2 - 5.900 \left(\frac{T}{T_B} \right)^3, \quad \dots (3)$$

where T_B = boiling point.

It is apparent that the two equations, (2) and (3), are identical if the quantity Q is assumed to be variable with temperature. Thus

$$\begin{aligned} A + \frac{Q}{2.3RT} &= 6.283 - 18.572 \left(\frac{T}{T_B} \right) + 17.505 \left(\frac{T}{T_B} \right)^2 - 5.900 \left(\frac{T}{T_B} \right)^3, \quad (4) \\ &= 6.283 - f \left(\frac{T}{T_B} \right), \end{aligned}$$

where
$$f \left(\frac{T}{T_B} \right) = 18.572 \left(\frac{T}{T_B} \right) - 17.505 \left(\frac{T}{T_B} \right)^2 + 5.900 \left(\frac{T}{T_B} \right)^3.$$

Over an interval of temperature, ΔT , it may be written,

$$\Delta \left(\frac{Q}{T} \right) = -2.3R \Delta f \left(\frac{T}{T_B} \right) \quad \dots \dots \dots (5)$$

That this equation applies to all temperatures under which the normal paraffins exist as liquids may be shown as follows:—

The energy of viscosity Q is related to the energy of vaporization simply:

$$Q = \frac{Ml}{n},$$

where M = molecular weight,

l = latent heat of vaporization/gm.,

n = constant equal to 4.13 for normal paraffins ⁽³⁾.

As l is zero at T_C , the absolute critical temperature, then, if equation (5) is true for all temperatures at which the normal paraffins are liquids, it follows that

$$\frac{Q_C}{T_C} = \frac{Q_1}{T_1} - 4.55 \left[f \left(\frac{T_C}{T_B} \right) - f \left(\frac{T_1}{T_B} \right) \right] = 0, \quad \dots \dots (6)$$

where Q_1 = energy of activation at temperature T_1 .

As an illustration the case of n -octane may be worked out in full. It is to be remembered that equation (3) was obtained from a curve passing through the points plotted for various members of the paraffin family, and that n -octane itself need not have been taken into account in its derivation. Similarly, $\frac{Q_1}{T_1}$ may be taken from vapour pressure data to

Thus it may be concluded that in $\log \mu = A + \frac{Q}{2.3RT}$:

(a) A is invariant with temperature,

(b) Q is variable with temperature.

Further,
$$Q = T \left[\frac{Q_1}{T_1} - 4.55 f \left(\frac{T}{T_B} \right) \right],$$

$$= 2.3RT \left[6.283 - f \left(\frac{T}{T_B} \right) - A \right], \quad \dots \dots (8)$$

as can be seen from equation (4). A study of the terms A and Q reveals three types of energy affecting viscosity of liquids.

The most successful attempt at evaluating A in equation (2) from theoretical derivation was made by Eyring and his collaborators⁽⁶⁾. These workers obtained

$$A = \log \left[1.09 \times 10^{-3} \frac{M^{1/2} T^{3/2}}{V^{2/3} \Delta E_{\text{vap.}}} \right], \quad \dots \dots (9)$$

where

V = specific volume,

$\Delta E_{\text{vap.}}$ = molal latent heat of vaporization.

The correlation of the theoretical and experimental values of A did not yield entirely satisfactory results, however. This discrepancy was attributed to certain approximations implicit in the derivation, notably to assumptions regarding the length of unit jump of the molecules between two positions of equilibrium and to the assumption that monomolecular and not polymolecular types of reactions are involved. The present findings render another explanation possible. As the method used by Eyring gave such good approximations to the results, let it be assumed that the two assumptions which were doubted be correct, and

the term $\Delta E_{\text{vap.}}$ be examined instead. It is shown that the term $\frac{Q}{2.3RT}$ vanishes at the critical temperature. Then, as μ_C is finite quantity it must be assumed that A or its antilog is finite too. In fact

$$\mu_C = \frac{1.09 \times 10^{-3} M^{1/2} T^{3/2}}{V^{2/3} \Delta E_{\text{vap.}}}.$$

As $\Delta E_{\text{vap.}}$ is zero at the critical temperature, it follows that the energy term ΔE is, in fact, not the latent heat of vaporization, but some other form of energy instead. This conclusion would be reached even if A or its antilogarithm were variable with temperature. It is assumed that Eyring's derivation of A is correct in all respects except with respect to the energy term. Thus, the energy term can be calculated for each temperature from equation (9) by substituting for the values of A as known experimentally, and for M, T and V. In this way, for instance, the value of the energy term in the case of *n*-octane at 25°C. is 15920 cal./mole. or twice the energy of vaporization. At the critical temperature of *n*-octane, and assuming the constancy of A, the energy term is 15030 or only 6 per cent. less than its value at 25°C. Thus, by assuming a new type of energy to be involved rather than the latent

heat of vaporization, the difficulty of attaining infinite viscosity at the critical temperature is eliminated. However, this energy term appears to be almost invariant with temperature and its significance presents difficulties in interpretation.

The other two types of energy are implicit in Q . Equation (8) may be written :—

$$Q = 2.3RT \left[(6.283 - A) - 18.572 \left(\frac{T}{T_B} \right) + 17.505 \left(\frac{T}{T_B} \right)^2 - 5.900 \left(\frac{T}{T_B} \right)^3 \right],$$

$$= 2.3RT \left[(6.283 - A) + 17.505 \left(\frac{T}{T_B} \right)^2 \right]$$

$$- 2.3RT \left[18.572 \left(\frac{T}{T_B} \right) + 5.900 \left(\frac{T}{T_B} \right)^3 \right].$$

Thus, Q is the difference between positive and negative types of energy.

$$Q = E_{\text{pos.}} - E_{\text{neg.}}, \quad (10)$$

where $E_{\text{pos.}} = 2.3RT \left[(6.283 - A) + 17.505 \left(\frac{T}{T_B} \right)^2 \right], \quad (11)$

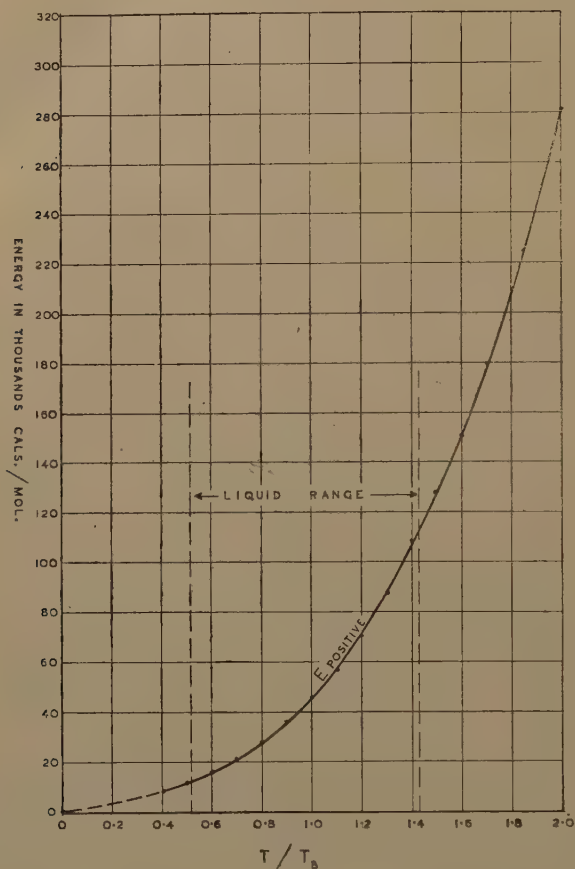
$$E_{\text{neg.}} = 2.3RT \left[18.572 \left(\frac{T}{T_B} \right) + 5.900 \left(\frac{T}{T_B} \right)^3 \right]. \quad (12)$$

These three terms were calculated for *n*-octane as an illustration and are shown in the table :—

T/T_B .	T .	$E_{\text{pos.}}$	$E_{\text{neg.}}$	Q .
0.50	199.0	11254	9075	2179
0.60	238.8	15597	13490	2107
0.70	278.6	21077	19042	2035
0.80	318.4	27907	25898	2099
0.90	358.2	36235	34248	1987
1.00	398.0	46281	44311	1970
1.10	437.8	58229	56331	1898
1.20	476.6	72127	70429	1698
1.30	516.4	88324	87175	1149
1.40	556.2	109125	108680	445
1.50	596.0	128645	129530	—885
1.60	635.8	152915	155915	—3000
1.70	675.6	180250	186200	—5950
2.00	796.0	282740	305440	—22700

In fig. 1 is shown a plot of $E_{\text{pos.}}$. The graph for $E_{\text{neg.}}$ is not much different except that it crosses $E_{\text{pos.}}$ curve at the critical temperature and climbs at a steeper rate. This can be better understood from fig. 2, which gives the variation of $Q = E_{\text{pos.}} - E_{\text{neg.}}$. It is seen that from fig. 2 the critical temperature could be correctly predicted from viscosity data alone. It also explains the apparent invariance of Q with temperature for small ranges in temperature below the boiling point. Further, it could be

Fig. 1.



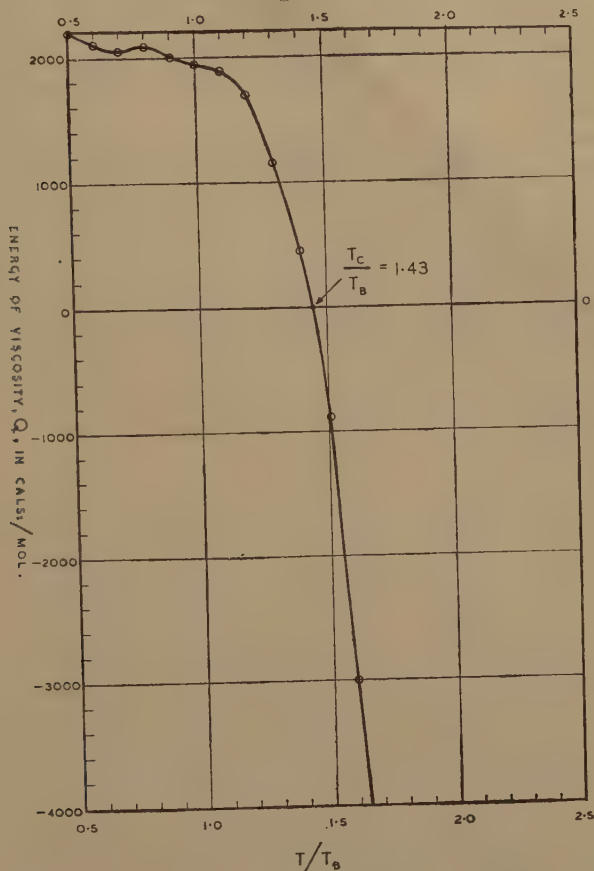
predicted that for a liquid maintained under critical pressure the viscosity would drop with a rise in temperature until a minimum is reached at the critical temperature when, due to the change in the sign of Q , it would rise in the gaseous state.

Conclusions.

Three terms of energies are found to determine the value of the viscosity of simple liquids. The first value appears to be invariant or only slightly affected by temperature. The two other terms both

increase rapidly but at different rates with temperature. They equate at the critical temperature and, as they possess opposite signs, cancel each other *. Probably similar results are obtained with latent heat of

Fig. 2.



* A tentative explanation of the three energies is as follows:—Each molecule is surrounded by fields of force, these being the resultants of attractive and repulsive forces varying with high but different powers of distance. At some point the two forces equate and the equipotential surface passing through all these zero points has been defined as the “shape” of the molecule⁽³⁾. Molecules tend to “touch” each other with these surfaces. It is tentatively suggested that at the critical temperature and pressure the molecules do so at these surfaces. The energy of separation from these equilibrium surfaces is given by the first energy described here, or by a function of this energy. At higher or lower temperatures and pressures molecules do not “touch” at these surfaces but they are either farther apart or nearer—*i. e.* “interfering” in each other’s zones of repulsive influences. Thus the energy of separation is increased or decreased by energy increments depending on the relative magnitudes of the repulsive and attractive forces under each condition. This hypothesis would explain, partly, the apparent invariance of the first energy and the variation with high powers of temperature of the second and third energies and their equating at T_c .

vaporization, as the energy of viscosity is a constant fraction of the heat of vaporization and both vanish at the critical temperature. Finally, it appears that there is an inherent increase in the viscosity of liquids when the temperature is raised, masked by a larger decrease. The opposite is probably true of gases. From these it appears that the mechanism of viscosity is probably the same in essentials for the two states of matter.

References.

- (1) Nissan, A. H., and Clark, L. V. W., 'Nature,' cxliii. p. 722 (1939).
- (2) Nissan, A. H., Phil. Mag. (ser. 7), xxxii. pp. 441-456 (1941).
- (3) Nissan, A. H., Clark, L. V. W., and Nash, A. W., J. Inst. Pet. xxvi. pp. 155-222 (1940).
- (4) Bridgman, P. W., Proc. Nat. Acad. Amer. xi. p. 603 (1925).
- (5) Newitt, "High Pressure Plant and Fluids at High Pressure." Oxford University Press, p. 408 (1940).
- (6) Eyring, H. *et al.*, J. Chem. Phys. iv. p. 283 (1936); *ibid.* v. p. 726 (1937).

Dept. of Oil Engineering and Refining,
The University,
Edgbaston,
Birmingham.

LV. *Temperature, Pressure, and Specific Volume Changes of a Gas under Dissociation and Reassociation Conditions.*

By Professor W. J. WALKER, University of the Witwatersrand,
Johannesburg, South Africa*.

[Received November 23, 1942.]

PRÉCIS.

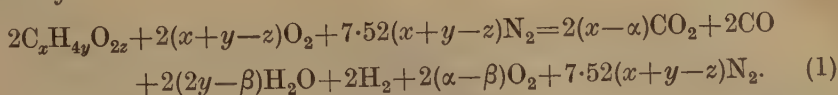
The following analysis provides a direct computation method of arriving at the temperature and pressure of explosion of a mixture of fuel and air, taking dissociation into account. The method applies also to the re-association correction required for temperature, pressure and specific volume changes during subsequent expansion.

IN the Report of the Empire Fuels Committee†, Tizard and Pye give a method of determining the products of combustion and the temperature at the end of constant volume combustion for an explosive mixture of fuel and air. The method is somewhat lengthy and involved, so much so that recourse is had to arriving at the final solution tentatively and not directly. The following method is general and gives a direct solution.

* Communicated by the Author.

† Proc. Inst. Aut. Engrs. vol. xviii. pt. 1 (1923).

The reaction equation for combustion, with dissociation, of any hydrocarbon or organic substance included in the general formula $C_xH_yO_z$ is given by



From this the ratio of the final to the initial volume, at constant pressure, is given by

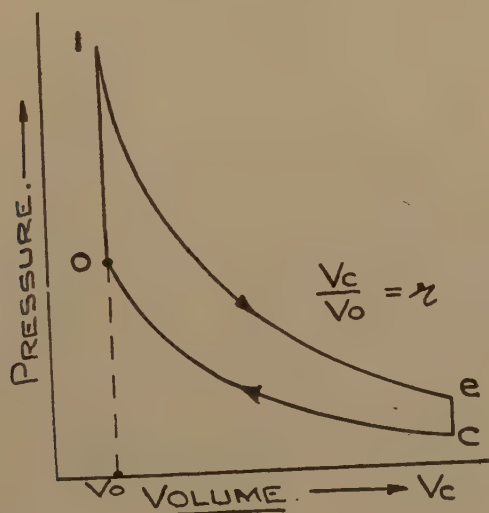
$$= \frac{x+1.21y-0.79z+0.21\alpha-0.105\beta}{x+y-z+0.21} \quad (2)$$

The relationship between the partial pressures of all the gases present in the products of combustion is given by

$$\frac{P_{CO} \times P_{H_2O}}{P_{CO_2} \times P_{H_2}} = K,$$

from which

$$\beta = \frac{2y\alpha}{K(x-\alpha) + \alpha} \quad (3)$$



Suppose the explosion to occur in an internal combustion engine following the cycle of the figure. To obtain a relation between α and T_1 , it will be necessary to know the maximum temperature T_{m1} , which would have been attained had there been no dissociation. This can readily be found by calculation or, alternatively, from a variable specific heat chart. Also, since it is permissible, to a close degree of approximation in expressions for ratio quantities, to neglect the effect of variable specific heat, the following expression is obtained :

$$\frac{T_1 - r^{\gamma-1}T_c}{T_{m1} - r^{\gamma-1}T_c} = \frac{C_1(x-\alpha) + C_2\alpha + C_3(2y-\beta)}{C_1x + 2C_3y}, \quad (4)$$

become so extremely sensitive, that attempts can now be made to solve the problem of the nature and mode of action of the mechanism which is at work.

It is customary to express the results of observation in terms of the effects of three luminous stimuli selected from the close neighbourhood of each of three suitable wave-lengths in the visible spectrum furnished by a standard, or specified, source of illumination. These wave-lengths are usually selected from the red, green and blue regions of the spectrum, and are called the Instrumental Fundamentals. It is now fully admitted that any colour whatsoever can be simulated (matched) by proper combination of three such stimuli. This was the basic statement of the so-called Trichromatic Theory.

But it is well known that the essential triplicity of phenomena of colour sensation can also be specified in terms of the three qualities of Intensity or Brightness, Hue or Tone, and Saturation or its converse—Whiteness. When two flames are made to agree in these three aspects, it is utterly impossible to distinguish their visual qualities. Helmholtz long ago pointed this out as a *proof* of trichromasy. And it was definite. It therefore becomes a matter of great importance to know the manner of dependence of these three qualities upon the three mutually independent fundamental stimuli: for, in the most recent physiological observations, this question is involved.

The Basic Laws and the Threshold Scale.

2. Mathematically the problem is simply one of expressing the basic laws which are found to agree sufficiently closely with experimental tests to be acceptable as at least good first approximations. The first law requiring expression is that of Colour Mixture; and the law now known as Newton's Law of Colour Mixture is a good first approximation to the truth. Deviations from it were actually used by Helmholtz for the explanation of more recondite phenomena and the prediction of others. The second law is the law connecting sensation with stimulus—the so-called Psychophysical Law. Fechner's law suits as a first approximation, the conditions for which have recently become more fully known. Higher approximations were also used by Helmholtz. The third law may be stated in different ways which require a fuller discussion here, as they are not so generally known to workers in consequence of their not being involved in most discussions of observed visual effects.

The name given to it is the Law of Coefficients. The most convenient way of discussing its meaning for the present purpose is through consideration of the psychophysical law in its simplest form, which gives the value of the sensation S due to a stimulus s as

$$S = k \log \frac{s}{s_0},$$

where k and s_0 are constants—the latter being subject to great variation

under slight change of experimental conditions. But its meaning, under specified conditions, is very easily recognizable: for, if $s_0 = s$, the sensation vanishes. So S is the least value of the stimulus which can give rise to the sensation; and it is therefore called the *Threshold* value of the stimulus—a *just perceptible step*. The actual value of the stimulus may therefore be expressed by the number of just perceptible steps in it. This number represents what I have elsewhere called the Sensation Value of the stimulus. It furnishes a scale in Colour Sensitivity quite analogous to a thermometry scale in the case of heat. And, in analogy also, the stimulus could be measured as energy, just as is done in the case of heat.

Further, the Colour Sensimetry scale has this advantage, over any ordinary scale of temperature, that it is quite independent of the unit in terms of which the energy of the stimulus is measured. The magnitude of the unit, that is the Threshold itself, can be altered by fatigue or defatigue (Prof. Frank Allen's "enhancement") of the eye. But that is exactly the feature that expresses the Sensitivity.

On the other hand, there is no *a priori* reason for any postulate regarding the absolute magnitude of successive steps in the process of building up the total number. This question is of the same nature as that of the uniform spacing of successive marked degrees on a thermometer. In a good mercury-glass thermometer the condition is fairly well satisfied: in a water-glass thermometer readings would be ambiguous through a considerable range. It has been customary to *postulate* uniformity of the scale.

Deduction of the Coefficient Law.

3. In accordance with that postulate, and assuming, for simplicity, the sufficient accuracy of the simplest form of the psychophysical law expressed above, we may deduce a necessary consequence of the approximate accuracy of the assumptions.

From the general formula

$$S = k (\log s - \log s_0)$$

we get

$$dS = k \left(\frac{ds}{s} - \frac{ds_0}{s_0} \right).$$

Now let fatigue by white illumination have produced a change of threshold large in comparison with ds_0 . So we have

$$dS = k \left(\frac{ds}{s} - \frac{ds'_0}{s'_0} \right).$$

And if we now put

$$dS' = k \left(\frac{ds'}{s'} - \frac{ds'_0}{s'_0} \right),$$

$$d(S' - S) = k \left(\frac{ds'}{s'} - \frac{ds}{s} \right),$$

we can solve, to the first order of small quantities, the problem of finding the new stimulus s' which is then necessary to maintain the state s_0' .

If, therefore, we make the same fractional change of stimulus in the two sets of observations, we arrive at the conclusion that there will be no change in sensation.

But this is a particular case of the Law of Coefficients as stated by v. Kries. It is here deduced from the psychophysical law under the assumption of the uniformity of the scale of just perceptible sensations.

Conversely, we may derive from the approximate truth of the psychophysical law, and that of the law of coefficients, the, at least approximate, truth of the law of thresholds.

Simplicity and predictive power are the two normal tests for the choice of a postulated law. Constancy of the Threshold is the simplest of all laws that it may obey.

Composition of Sensation Values.

4. The sensation value, being the number of just perceptible steps in a stimulus, and the scale of just perceptible steps being accepted as uniform, it follows that the total sensation value of a stimulus which involves all three fundamentals is the sum of the three individual values. If c and c_0 represent respectively the values of the total stimulus and the threshold stimulus in a coloured (or white) illumination of the retina, we therefore obtain the relation

$$\frac{c}{c_0} = \frac{r}{r_0} + \frac{g}{g_0} + \frac{b}{b_0},$$

where r, g, b are the values of the experimental fundamentals employed. This can be put in the form

$$\frac{1}{c_0} = \frac{1}{r_0}f(r) + \frac{1}{g_0}f(g) + \frac{1}{b_0}f(b),$$

where the f 's are the fractional red, green and blue stimulations supplied by the apparatus.

In the first form it gives the Law of Sensation Values. In the second form it states the Law of Thresholds: or, if we use the numerical values of the threshold reciprocals, it gives the Law of Coefficients.

It has direct application to the recent noteworthy work of Dr. W. D. Wright (*e.g.*, Proc. R. S., B, vol. cxv. pp. 74-76) and Prof. Gravit ('Nature,' 2 Jan. 1943) in their attempts to investigate the nature of the mechanism and the processes which are active in colour vision.

The difficulties which have been thought to arise in connexion with the Coefficient Law, and the consequent tendency to return to the vagueness of the conception of Modulation are brought about by non-recognition of the uniformity of the psychological scale of sensation. The difficulties vanish, and the vagueness disappears, when the soundness of that law is recognized. The three laws—Psychophysical, Coefficient,

and Uniformity of the Threshold Scale—are not independent. Any one is a consequence of the other two. The third is by far the most simple in statement, and the first follows it in order of simplicity. This does *not* mean that the Coefficient Law is to be relegated to an inferior position. On the contrary, its elucidation in all its varied aspects and consequences is very pressing. The recent work, through great improvement of both physical and physiological instrumental resources, is of the highest importance.

But, in one aspect of that work, discussion in terms of modulation seems to be apt to involve incomplete recognition of the other three decisive aspects of chromatic triplicity—brightness, hue and saturation—which are *necessary consequences* of the former three—independent red, green and blue absolute fundamental stimuli and sensations.

Quantity and Quality in Colour Vision.

5. When numerical magnitude alone is involved in any phenomenon, that phenomenon is said to be a scalar one. That is to say, one number, which can be registered on a scale, is sufficient to specify it. But, if quality is also present, one number is no longer sufficient. Two, three, or more are needed in accordance with the multiplicity of the quality. In colour vision, the number now agreed upon as necessary and sufficient is three.

Phenomena of colour vision can therefore be represented by a *vector* in tridimensional space. This was one of the important ideas introduced by Helmholtz—quality being analogous to distinctive direction in space. The *manner* of its representation, however, is one to be settled in accordance with experimental data.

Now the *Intensity* of light corresponds to the sensation of Brightness, which is characteristic of any light whatever may be its quality as to colour. So it is a purely scalar characteristic. Therefore, colour is specifiable by measuring the magnitudes of the three chosen fundamental colours, which are present in a compound light, along three selected mutually perpendicular lines. This is well known to all students of the subject. But there is a fundamental difficulty with which Helmholtz was faced. Any function of the three component brightnesses might be measured along the axes, and difference of quality would still be shown by the vector direction—more or less simply in accordance with the choice of the function. The magnitude of the resultant vector will vary accordingly, and that magnitude must represent in some way the resultant intensity. The usual vector representation gives the resultant as the square root of the summed squares. This was naturally the one tried by Helmholtz.

The question remains “Is it in conflict with any of the known laws of vision”? Now Newton’s Law of Colour Mixture, which is a good first approximation, asserts that the total intensity of superposed components is the arithmetical sum of the component intensities. At

that time Helmholtz had grave doubts as to the measured magnitudes of the sensations evoked in the case of coloured lights. It is now known that the measurements of intensity, made by methods free from the colour difficulty, support the application of Newton's Law (known as Abney's Law in the case of the three variables). It was pointed out by W. Pauli that a suitable function of the intensity to use in the vector construction is the square root of the intensity. *This makes the vector construction lead to Abney's Law.* It is appropriate to the propagation of light, since light acts as if it were propagated through a medium which obeys Hooke's Law.

But it is the propagation of *energy* with which we are concerned, and that is generally carried out by means of all three of the *independent* activities proceeding between the retina and the cortex.

Since these activities seem to be of electrical type, the rate of supply is proportional to the square of the current or of the potential: so that Pauli's construction can be employed.

Yet r , g and b being measured as energy, and since it is these component energies which are *independently* propagated, we may still legitimately raise the question "Is not Helmholtz's vector space the correct one to employ, with his resultant as the proper one"? In other words, "Can we get reconciliation between Abney's Law *otherwise* than through Pauli's plan"? The answer is given by means of the Law of Sensation Values whenever we recognize the uniformity of the Threshold Scale of Energy.

The Expressions for Brightness, Hue, and Saturation.

6. These expressions can very readily be given, either by a geometrical construction or as analytical functions of the fundamental stimuli r , g , b . And the expressions, in either way, are very simple. The investigation of the difference of colour sensation between two-colour vectors was first worked out by Schrödinger. It is proportional to the angle between the two colour vectors. In Pauli's space it is equal to twice the angle between the two vectors: in ordinary space it is equal to the angle itself. Similarly, in Pauli's space, the intensities are proportional to the squares of the vector lengths. This agrees with Abney's Law, since the co-ordinate quantities are the square roots of r , g , b . In ordinary space, with co-ordinates r , g , b , the resultant intensity is the one used by Helmholtz—the square root of the sum of the squares of r , g , b .

A brief treatment of the whole problem, in Pauli's and also in ordinary space, is given in 'Nature,' Nov. 23, 1929, by the quaternionic analysis, which makes great condensation possible.

The further analysis of the colour difference into difference of hue and difference of saturation leads, *in Pauli's space*, to the result that, if the white vector be taken as a polar axis, the difference in *saturation* which appears as we proceed from a vector denoted by the suffix 1 to

another denoted by the suffix 2 is twice the difference of the polar distances.

And the difference in *hue* which arises in the same procedure is given by the product of the difference in longitude of the two-colour vectors into twice the value of the initial polar distance expressed in terms of a right angle as the unit. The difference in hue vanishes at the white point not merely because the colour stimulus goes below the threshold there, but in its own essence from the vanishing of the polar distance.

Domination and Modulation.

7. The reason why I have given these details in the preceding section is primarily to secure recognition of the position there—*anent* even by workers who may have no adequate knowledge of the steps by which the deductions are arrived at.

Another reason is consequent to that one. It is to ensure that no inconsistent conclusion regarding change of intensity or of colour effects may be drawn through the use of a different terminology.

The final object is to make clear the relationships of the views regarding Domination and Modulation, on the one hand, and of Fatigue along with Defatigue (or Enhancement) on the other. I do not feel that there is gain, but, on the contrary, loss through the departure from Helmholtz's terminology.

The word "Modulation" was introduced by Prof. Southall, editor of the third (American) edition of Helmholtz's 'Physiological Optics'—see footnote, pp. 139, 140, vol. ii. of that edition, which is a translation of the third German edition. Professor Southall gives it as a more suitable equivalent of the word "unstimmungen" than such terms as "mutation," "conversion," etc. It is genuinely better. The word "umstimmungen" is used by v. Kries in his Appendix to Helmholtz's text. Its first use occurs in Nagel's Appendix to the original text in vol. ii. p. 350.

The word "fatigue" means simply "rise of threshold"; and the words "recovery," "defatigue" or "enhancement" imply merely "decrease of threshold." All internal changes which influence activity alter the threshold, and are represented in the mathematical formulation. Recent work, referred to above, has its origin in exceedingly interesting attempts to determine the *cause* of the changes of activity; and so are contributions to knowledge regarding the *mechanism*, which is outside the aim of the pure theory. But they might quite conceivably give rise to a modification of the presently accepted theoretical laws.

In my view departure from Helmholtz's phraseology is entirely unnecessary. But there can be no intrinsic objection to the use of the terms Dominator and Modulator except in so far as they may lead to the idea that they represent *fundamentally* different influences. They represent largely similar variations of threshold, which must occur whenever incident light energy, or transmitted energy, is transformed.

Their investigation gives a most welcome way of investigating the peculiarities in the form of the so-called "sensation" curves—quite apart from the special aim at finding the *cause* of the peculiarities.

It may happen to be found that the "dominator" activity is always due to one type of cause of change different from those made evident in the "modulator" activities. In that case the two terms would acquire greater value than that of correspondence to width or narrowness of absorption bands in the internal electric spectrum.

The Physical Basis of Helmholtz's Vector Construction.

8. The features of the problem, discussed in Sec. 3, enable us to realize the exact position of the facts concerning Helmholtz's vector procedure in colour vision. There is a very glaring, if only seeming, discrepancy between Helmholtz's expression for the magnitude of the resultant colour vector and the expression apparently required by Newton's Law of Colour Mixture as corroborated by Abney. That verification holds to a sufficiently good approximation to safeguard the law in this matter. But it seems that Helmholtz's Colour Vector Law is equally safeguarded by the fact that the triply *independent* activities, which are propagated cortex-wards, correspond to component amounts of *energy* which, in the vector diagram, *must* be represented by mutually perpendicular lines in order to secure the independence. How then is reconciliation of the apparently discordant views to be effected? Only by providing for obedience to Abney's Law *otherwise* than by Pauli's ingenious construction.

The basis for Helmholtz's construction must lie in a region outside the sway of the Fechner-Helmholtz rulings. It must provide a particular scheme of energy-measurement which does not merely *permit* but actually *secures* the Helmholtz-Newton agreement. The provision is that we adopt, for the energy measurement, the uniform scale given by the equality of successive just-perceptible steps in the Sensation Value of Brightness.

The small term, neglected in the last equation of sec. 3, is

$$k \left(\frac{ds_0'}{s_0'} ds_0' - \frac{ds_0}{s_0} ds_0 \right).$$

The condition for its complete vanishing is that the difference of the square roots of s_0' and s_0 shall be constant. In the normal condition of steady rise of the threshold accompanying a steady rise of stimulus (*i. e.* with no independently imposed fatigue) this condition is satisfied, since

$$\frac{ds_0'}{ds'} = \frac{ds_0}{ds} = 1$$

for the uniform scale. So the arithmetical mean of s_0' and s_0 , with first-order difference of the two only, is equal to the geometrical mean. This brings Helmholtz's scheme into agreement with the Law of Newton.

LVII. *The Fundamental Equation of Quantum Mechanics.*

By H. T. FLINT, Reader in Physics,
University of London, King's College*.

[Received January 26, 1943.]

SCHROEDINGER's equation for a particle of charge e and mass m_0 in an electromagnetic field is

$$\sum_{m=1, 2, 3} \left(\Pi_m - \frac{e}{c} \phi_m \right)^2 \psi - \frac{1}{c^2} \left(\frac{\hbar}{2\pi i} \frac{\partial}{\partial t} + e\phi \right)^2 \psi + m_0^2 c^2 \psi = 0, \quad (1)$$

where Π_m denotes the momentum operator $\frac{\hbar}{2\pi i} \frac{\partial}{\partial x^m}$ and the other symbols have their usual meanings.

This equation is defective in that it takes no account of the magnetic moment of the particle. This defect has been remedied, as is well known, by Dirac, who introduced an equation of the first order as the fundamental equation of the quantum theory.

It is proposed to show how the defect may be remedied in quite a different and perhaps more natural way. The principle of the method has something in common with one introduced by Schroedinger⁽¹⁾, but there are important differences.

The present method requires the use of five co-ordinates and, if these are regarded as co-ordinates in a system of geometry in which the line element is

$$d\sigma^2 = \gamma_{\mu\nu} dx^\mu dx^\nu, \quad \dots \quad (2)$$

the invariance of the principle of relativity is automatically satisfied. The four Dirac matrices $(\alpha_x, \alpha_y, \alpha_z, \alpha_0)$ will, in the first place, be replaced by five of which γ^μ is typical. These satisfy Tetrad's relations extended to this five-dimensional case, viz.:

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\gamma^{\mu\nu}. \quad \dots \quad (3)$$

The following relations due to Kaluza, which he introduced for the purpose of unification of gravitation and electromagnetism, are adopted for the present purpose:

$$\left. \begin{aligned} \gamma^{mn} &= g^{mn}, & \gamma^{m5} &= -\alpha \phi^m, & \gamma^{55} &= 1 + \alpha^2 \phi_m \phi^m, \\ \gamma_{mn} &= g_{mn} + \alpha^2 \phi_m \phi_n, & \gamma_{m5} &= \alpha \phi_m \end{aligned} \right\}, \quad \dots \quad (4)$$

where m and n denote the integers 1, 2, 3, 4 and ϕ_m is a component of the electromagnetic potential.

In addition it is assumed that $\gamma_{55} = 1$ and $\alpha = e/m_0 c^2$ ⁽²⁾.

* Communicated by the Author.

In operations of differentiation concerning matrices like γ^μ and single-rowed or column matrices such as ψ occurring in the quantum theory, it is necessary to take account of the concept of parallel displacement ^{(1), (3)}. Such a quantity, ψ , which may also be described as a vector with four components, is differentiated according to the rule,

$$\frac{d\psi}{dx^\nu} = \frac{\partial\psi}{\partial x^\nu} - \Delta_\nu\psi. \quad (5)$$

In the case of a four-rowed matrix, F ,

$$\frac{dF}{dx^\nu} = \frac{\partial F}{\partial x^\nu} + F\Delta_\nu - \Delta_\nu F. \quad (6)$$

There are five operators (Δ_ν) which are analogous to the bracket expressions employed in the theory of relativity in covariant differentiation. If the matrix has the character of a covariant or contravariant vector or tensor, the principles of covariant differentiation must be applied. In the case of a matrix (F^μ)

$$\frac{dF^\mu}{dx^\nu} = \frac{\partial F^\mu}{\partial x^\nu} + \Delta_{\rho\nu}^\mu F^\rho + F^\mu \Delta_\nu - \Delta_\nu F^\mu, \quad (7)$$

where $\Delta_{\rho\nu}^\mu$ is the Christoffel bracket expression in five-dimensional geometry.

The quantities (γ^μ , γ_μ) satisfy relations, of which the following is typical:

$$\frac{\partial\gamma_\mu}{\partial x^\nu} = \gamma_\rho \Delta_{\mu\nu}^\rho + \Delta_\nu \gamma_\mu - \gamma_\mu \Delta_\nu. \quad (8)$$

This relation corresponds to the identity

$$\frac{\partial g_{mn}}{\partial x^i} = \Gamma_{mi}^r g_{rn} + \Gamma_{ni}^r g_{mr},$$

and the values of Δ_ν can be obtained from it.

It is necessary to express results in terms of the four-dimensional continuum so that they can be easily interpreted. To do this it is to be observed that any five-dimensional vector can be replaced by a four-dimensional vector and a scalar. This scalar is the fifth component of the covariant five-vector whereas the four-vector is obtained by taking the first four components of the contravariant five-vector.

In order to preserve a distinction between the two representations different letters will be used. This is illustrated in the case of the vector (γ^μ) by denoting the four-dimensional counterparts by (β^m) and β . The relations are

$$\beta^m = \gamma^m, \quad \beta = \gamma_5.$$

It follows that

$$\beta_m = \gamma_m - \alpha \phi_m \gamma_5, \quad \gamma_5 = \gamma_5 - \alpha \phi_m \beta^m.$$

From the relations (3) Tetrode's relations follow,

$$\beta_m \beta_n + \beta_n \beta_m = 2g_{mn},$$

and in addition

$$\begin{aligned}\gamma_5 \beta_m + \beta_m \gamma_5 &= 0, \\ \gamma_5^2 &= \beta^2 = 1.\end{aligned}$$

Since it is not necessary to take account of the gravitational field in the case of applications of the quantum theory, it is sufficient to write $g_{mn} = \delta_{mn}$ and to make no distinction between contravariant and covariant quantities in four dimensions. The values of the operators (Δ_v) have been shown to be ⁽³⁾

$$\left. \begin{aligned} \Delta_n &= \frac{1}{2} \alpha (\alpha \phi_n H_{lp} \beta_p \beta_l + H_{ln} \gamma_5 \beta_l), \\ \Delta_5 &= \frac{1}{2} \alpha H_{lp} \beta_p \beta_l, \end{aligned} \right\}, \quad \dots \quad (9)$$

where

$$H_{lp} = \frac{\partial \phi_l}{\partial x^p} - \frac{\partial \phi_p}{\partial x^l}.$$

Dirac's equation is usually written in the form

$$\left(\frac{\hbar}{2\pi i c} \frac{\partial}{\partial t} + \frac{e}{c} \phi \right) \psi = \Sigma \alpha_k \left(\Pi_k - \frac{e}{c} \phi_k \right) \psi + m_0 c \alpha_0 \psi.$$

The form introduced in the present theory is ⁽⁴⁾

$$\Sigma_{m=1, 2, 3, 4} \beta^m \left(\Pi_m - \frac{e}{c} \phi_m \right) \psi + m_0 c \gamma_5 \psi = 0.$$

Thus

$$\begin{aligned}\alpha_x &= \alpha_1 = -i \beta_4 \beta_1, \\ \alpha_y &= \alpha_2 = -i \beta_4 \beta_2, & \alpha_0 &= -i \beta_4 \gamma_5, \\ \alpha_z &= \alpha_3 = -i \beta_4 \beta_3,\end{aligned}$$

in which the distinction between β^m and β_m has been dropped.

The natural generalization of the equation

$$p_x^2 + p_y^2 + p_z^2 + p_4^2 + m_0^2 c^2 = 0,$$

is

$$\gamma^{\mu\nu} \Pi_\mu \Pi_\nu = 0,$$

where Π_μ is a momentum component in the five-dimensional continuum,

i. e.,
$$\Pi_\mu = m_0 \gamma_{\mu\nu} \frac{dx^\nu}{dt},$$

and in the case of Π_5 it is part of the theory that $\Pi_5 = m_0 c$. Thus, replacing the equation by its quantum analogue, it is to be expected that the second-order quantum equation will be

$$\gamma^{\mu\nu} \Pi_\mu \Pi_\nu \psi = 0, \quad \dots \quad (10)$$

where Π_μ now denotes the operator.

In the particular case of Π_5 we require

$$\Pi_5 \psi = m_0 c \psi, \quad \dots \quad (11)$$

in other words it is implied that when x^5 occurs as in the quantity ψ , it does so in the form of a factor $\exp(2\pi i m_0 c x^5 / \hbar)$. Equation (10) is now identical with (1) if the relations (4) and (11) are applied.

According to the view expressed here, the reasons why equation (10)

like equation (1) is incomplete is because the operation of differentiation is not correctly applied. Equation (10) is not the true quantum analogue on account of the character of ψ as a single-column matrix.

The correct equation should be

$$\gamma^{\mu\nu} \left(\Pi_\mu - \frac{\hbar}{2\pi i} \Delta_\mu \right) \left(\Pi_\nu - \frac{\hbar}{2\pi i} \Delta_\nu \right) \psi = 0. \quad (12)$$

This equation satisfies not only the requirements of relativistic invariance but also those of the similarity transformation⁽¹⁾. The latter are automatically satisfied in the present case because due regard has been paid to the parallel displacement of operators.

Some explanation should be added on account of relativistic invariance, for the correct relativistic form of equation (12) is

$$\text{div} (\text{grad } \psi) = 0,$$

in which additional terms occur. These vanish in the present case where no gravitational field exists and where γ , the determinant of the $\gamma_{\mu\nu}$, is placed equal to unity. This latter condition is permissible and corresponds to a similar one in the theory of relativity.

In examining this equation it will be supposed that the energy, W , of the particle is small with respect to $m_0 c^2$. The first term is equal to

$$\Sigma \left(\Pi_m - \frac{e}{c} \phi_m \right)^2 \psi + m_0^2 c^2 \psi. \quad (13)$$

The term in this expression in which $m=4$ becomes, under the condition of this approximation, $(2m_0 e \phi - 2m_0 W - m_0^2 c^2)$.

In the expression for Δ_μ the product $\alpha \phi_\mu$ occurs. In the case $\mu=4$ the magnitude of this factor is $e\phi/m_0 c^2$ and can be neglected in this approximation. It will be assumed that the same condition applies to $\alpha \phi_\mu$ for all the values of μ . It is thus sufficient to write from the expressions (9)

$$\Delta_n = \frac{1}{4} \alpha H_{lm} \gamma_5 \beta_l,$$

and we shall omit squares of this quantity occurring in the equation.

Thus the second-order equation is simplified to

$$\gamma^{\mu\nu} \left(\Pi_\mu \Pi_\nu - \frac{\hbar}{2\pi i} \Delta_\mu \Pi_\nu - \frac{\hbar}{2\pi i} \Pi_\mu \Delta_\nu \right) \psi = 0. \quad (14)$$

The second term can be put into a simple form

$$\begin{aligned} \gamma^{\mu\nu} \Delta_\mu \Pi_\nu \psi &= \gamma^{mn} \Delta_m \Pi_n \psi + \gamma^{m5} \Delta_m \Pi_5 \psi + \gamma^{5m} \Delta_5 \Pi_m \psi + \gamma^{55} \Delta_5 \Pi_5 \psi \\ &= \Delta_m \Pi_m \psi - m_0 c \alpha \phi_m \Delta_m \psi - \alpha \phi_m \Delta_5 \Pi_m \psi + (1 + \alpha^2 \phi_m \phi^m) m_0 c \Delta_5 \psi. \end{aligned}$$

With the same approximation as before this becomes

$$\frac{1}{4} \alpha H_{lm} \gamma_5 \beta_l \Pi_m \psi + m_0 c \Delta_5 \psi.$$

Similarly the third term becomes

$$\frac{1}{4} \alpha \Pi_m (H_{lm} \gamma_5 \beta_l \psi) + m_0 c \Delta_5 \psi.$$

These taken together give

$$\frac{1}{2}\alpha H_{lm}\gamma_5\beta_l\Pi_m\psi + \frac{1}{4}\alpha(\Pi_m H_{lm})\gamma_5\beta_l\psi + 2m_0c\Delta_5\psi.$$

The second term in this expression contains the current density \mathbf{S} for

$$\frac{\partial H_{lm}}{\partial x^m} = S_l,$$

where summation over m is implied, as in all cases where the same suffix appears twice.

This will be placed equal to zero, since the charged particle is regarded as a test particle placed in the field.

In order to examine the nature of the first term let the operator Π_m be replaced by the momentum, $m\frac{dx^m}{dt}$, which it represents.

The quantity $H_{lm}\frac{dx^m}{dt}$ which arises is $-cF_l$, where F_l is a component of the Lorentz force on the particle ($l=1, 2, 3$). When $l=4$,

$$H_{4m}\frac{dx^m}{dt} = -\epsilon \mathbf{E} \cdot \mathbf{v},$$

where \mathbf{E} is the electric intensity of the field and \mathbf{v} the velocity of the particle.

The matrix product $\gamma_5\beta_l = \alpha_0\alpha_l$ ($l \neq 4$) and $\gamma_5\beta_4 = -\alpha_0$. Thus, writing $\eta_l = \alpha_0\alpha_l$, the term may be expressed in the form

$$-\frac{1}{2}\alpha m(c\mathbf{F} \cdot \boldsymbol{\eta} + \mathbf{E} \cdot \mathbf{v}\alpha_0).$$

Equation (14) now becomes

$$\begin{aligned} W\psi = & \left[e\phi + \sum_{m=1,2,3} \frac{1}{2m_0} \left(\Pi_m - \frac{e}{c}\phi_m \right)^2 \right. \\ & \left. - \frac{\hbar c}{2\pi i} \Delta_5 + \frac{\hbar \alpha m}{2\pi i 4m_0} (c\mathbf{F} \cdot \boldsymbol{\eta} + \mathbf{E} \cdot \mathbf{v}\alpha_0) \right] \psi. \quad (15) \end{aligned}$$

The term in Δ_5 can be put into a familiar form :

$$\begin{aligned} \Delta_5 = & \frac{1}{4}\alpha(H_{23}\beta_3\beta_2 + H_{32}\beta_2\beta_3 + \dots) = \frac{1}{2}\alpha(H_{23}\beta_3\beta_2 + \dots + H_{43}\beta_3\beta_4) \\ = & \frac{1}{2}\alpha(H_x\alpha_3\alpha_2 + \dots + E_z\alpha_3). \end{aligned}$$

Let the matrix $\boldsymbol{\sigma}$ be introduced such that

$$\sigma_1 = \sigma_x = \alpha_3\alpha_2, \text{ etc.}$$

Then

$$\Delta_5 = -\frac{ie}{2m_0c}(\mathbf{H} \cdot \boldsymbol{\sigma}) - \frac{e}{2m_0c^2}(\mathbf{E} \cdot \boldsymbol{\alpha}),$$

and the corresponding contribution to the energy is

$$-\frac{\hbar c}{2\pi i} \Delta_5 = \frac{\hbar e}{4\pi m_0c}(\mathbf{E} \cdot \boldsymbol{\sigma}) + \frac{\hbar e}{4\pi m_0c}(\mathbf{E} \cdot \boldsymbol{\alpha}).$$

This is the well-known expression for the additional energy characteristic of Dirac's theory of the electron.

The form of the second-order equation in this notation suggests that it may be regarded as a fundamental equation of the theory. But,

bearing in mind Dirac's method of approach to the first order equation, for reasons which, however, do not apply to equation (12), we naturally examine the equation

$$\gamma^\mu \left(\Pi_\mu - \frac{\hbar}{2\pi i} \Delta_\mu \right) \psi = 0. \quad (16)$$

This reduces at once to

$$\gamma^\mu \Pi_\mu \psi = 0, \quad (17)$$

since

$$\gamma^\mu \Delta_\mu \equiv 0.$$

Equation (17) is Dirac's equation⁽³⁾.

This procedure is without significance in the present case. The first-order equation gets its interpretation from quite a different consideration, for it arises as a gauging equation in a system of matrix geometry and metrics⁽⁵⁾.

Note added May 25th, 1943.—The additional energy terms can be reduced to an interesting form.

The term $\frac{1}{4} \alpha \Pi_m H_{lm} \gamma_5 \beta_l$ gives rise to an energy term

$$\frac{1}{8} \left(\frac{\hbar}{2\pi} \right)^2 \frac{a}{m_0} \left(\frac{4\pi}{c} \rho v_k \gamma_5 \beta_k + 4\pi i \rho v_4 \gamma_5 \beta_4 \right),$$

where the current density has been expressed in the form

$$S_k = 4\pi \rho v_k / c \text{ for } k=1, 2, 3,$$

where v_k is a velocity component and $S_4 = 4\pi i \rho$.

In general this is not equal to zero and, regarding the expression as an entity, it is found to have the characteristic values

$$\lambda = \pm \frac{1}{8\pi} \left(\frac{\hbar}{m_0} \right)^2 \frac{e}{c^2} \rho \sqrt{1 - \frac{v^2}{c^2}}.$$

These are real and can be regarded as arising from an interaction between the electron and the charge through which it is moving.

The term $\frac{1}{4} \alpha \Pi_m (H_{lm} \gamma_5 \beta_l \psi)$ requires more detailed consideration than is given above where $\Pi_m \gamma_5$ and $\Pi_m \beta_l$ were taken to be zero. They must be considered by means of equation (8) when it is found that they do not contribute to the final result.

The remaining term is

$$- \frac{\hbar}{2\pi i} \frac{a}{4m_0} H_{ln} \gamma_4 \beta_l \left(\Pi_n - \frac{e}{c} \phi_n \right)$$

which can be written in the form

$$\frac{\hbar}{8\pi} \frac{a}{m_0} H_{ln} c \gamma_5 \beta_l p_n,$$

where p_n is the momentum operator $(\Pi_n - e\phi_n/c)$.

If dp_k/dt be understood as a differentiation of the operator p_k in the sense appropriate to the quantum operator theory,

$$dp_k/dt = e(E + \alpha \times H)_k,$$

where

$$\alpha = (\alpha_x, \alpha_y, \alpha_z).$$

$(E = \alpha \times H)$ is the quantum analogue of the Lorentz force. This holds for $k=1, 2, 3$, and in addition,

$$dp_4/dt = ieE_k\alpha_k.$$

It follows that

$$eH_{in}\gamma_5\beta ip_n = \alpha_0 \frac{dp_n}{dt} p_n, \quad (n=1, 2, 3, 4),$$

and this additional term now takes the form $\frac{h}{8\pi m_0 c^2} \alpha_0 \frac{dp_n}{dt} p_n$.

References.

- (1) Schroedinger, *Sitzb. d. Preuss. Akad.* xi. (1932).
- (2) Flint, *Phil. Mag.* 7, xxix. p. 330 (1940).
- (3) Flint, *Phil. Mag.* 7, xxxiii. p. 369 (1942).
- (4) Flint, *Proc. Phys. Soc.* 1. p. 899 (1938).
- (5) Flint, *Proc. Roy. Soc. A*, cl. p. 432 (1935); *Proc. Phys. Soc.* xlviii. p. 433 (1936).

LVIII. A Steady State Problem in Heat Conduction.

By ARNOLD N. LOWAN*.

[Received November 5, 1942.]

CONSIDER a semi-infinite cylindrical annulus, bounded by the plane $z=0$ and the cylindrical surfaces $r=a$ and $r=b$.

If the surfaces $z=0$ and $r=a$ are kept at 0°C ., and the surface $r=b$ is kept at temperature $T=f(z)$, what is the distribution of temperature in the annulus?

The temperature $T(r, z)$ must satisfy the following differential equation and boundary conditions:

$$\frac{\partial^2 T}{\partial r^2} + \frac{1}{r} \frac{\partial T}{\partial r} + \frac{\partial^2 T}{\partial z^2} = 0, \quad \dots \dots \dots (1)$$

$$T=0 \quad \text{for } z=0, \quad \dots \dots \dots (2)$$

$$T=0 \quad \text{for } r=a, \quad \dots \dots \dots (3)$$

$$T=f(z) \quad \text{for } r=b, \quad \dots \dots \dots (4)$$

$$T \text{ is finite for } z \rightarrow \infty. \quad \dots \dots \dots (5)$$

Let $T(r, z) = R(r)Z(z). \quad \dots \dots \dots (6)$

* Communicated by the Author.

Then (1) becomes

$$Z\left(\frac{\partial^2 R}{\partial r^2} + \frac{1}{r} \frac{\partial R}{\partial r}\right) + R \frac{\partial^2 Z}{\partial z^2} = 0,$$

or
$$\frac{1}{R}\left(\frac{\partial^2 R}{\partial r^2} + \frac{1}{r} \frac{\partial R}{\partial r}\right) = -\frac{1}{Z} \frac{\partial^2 Z}{\partial z^2} = \lambda^2, \text{ (say),}$$

whence
$$\frac{\partial^2 Z}{\partial z^2} + \lambda^2 Z = 0 \quad \dots \dots \dots (7)$$

and
$$\frac{\partial^2 R}{\partial r^2} + \frac{1}{r} \frac{\partial R}{\partial r} - \lambda^2 R = 0. \quad \dots \dots \dots (8)$$

A particular solution of (7) satisfying the boundary conditions (2) and (5) is

$$Z = \sin(\lambda z). \quad \dots \dots \dots (9)$$

The general solution of (8) is of the form

$$R = AJ_0(\lambda r i) + BY_0(\lambda r i), \quad \dots \dots \dots (10)$$

where J_0 and Y_0 are the Bessel Functions of the first and second kind respectively, and A and B are arbitrary constants.

In order to satisfy the boundary condition (3), we must have

$$AJ_0(\lambda a i) + BY_0(\lambda a i) = 0,$$

whence
$$B = -\frac{AJ_0(\lambda a i)}{Y_0(\lambda a i)},$$

and solution (10) becomes

$$R = A \left[J_0(\lambda r i) - \frac{J_0(\lambda a i)}{Y_0(\lambda a i)} Y_0(\lambda r i) \right]. \quad \dots \dots \dots (11)$$

It is convenient to choose the constant A in the form

$$\frac{Y_0(\lambda a i)}{J_0(\lambda b i)Y_0(\lambda a i) - J_0(\lambda a i)Y_0(\lambda b i)} C,$$

where C is another integration constant. The solution (11) thus becomes

$$R = C \frac{J_0(\lambda r i)Y_0(\lambda a i) - J_0(\lambda a i)Y_0(\lambda r i)}{J_0(\lambda b i)Y_0(\lambda a i) - J_0(\lambda a i)Y_0(\lambda b i)}. \quad \dots \dots \dots (12)$$

But
$$J_0(\beta i) - I_0(\beta).$$

Further *,
$$Y_0(x) = -K_0(-ix) + J_0(x) \left\{ \text{Log } 2 - \gamma + \frac{\pi i}{2} \right\}, \quad \dots \dots \dots (13)$$

where γ is Euler's constant. If in (13) we replace x by xi , we obtain

$$Y_0(xi) = -K_0(x) + I_0(x) \left\{ \text{Log } 2 - \gamma + \frac{\pi i}{2} \right\}. \quad \dots \dots \dots (14)$$

In view of (14), solution (12) becomes, after some simple transformations,

$$\begin{aligned} R &= C \frac{I_0(\lambda r)K_0(\lambda a) - I_0(\lambda a)K_0(\lambda r)}{I_0(\lambda b)K_0(\lambda a) - I_0(\lambda a)K_0(\lambda b)} \\ &= CF(a, b, r, \lambda), \text{ (say).} \quad \dots \dots \dots (15) \end{aligned}$$

* See, for instance, Gray, Mathews and MacRobert, 'A Treatise on Bessel Functions.'

In view of (9) and (15), solution (6) becomes

$$T = CF(a, b, r, \lambda) \sin(\lambda z).$$

The last expression is a solution of the given differential equation, satisfying the boundary conditions (2), (3), and (5). It is clear that the superposition of an infinite number of solutions of this form yields a function satisfying the same differential equation and boundary conditions.

In view of the well-known double integral representation

$$f(z) = \frac{2}{\pi} \int_0^\infty \int_0^\infty f(\alpha) \sin(\lambda z) \sin(\lambda \alpha) d\lambda d\alpha,$$

it follows that the expression

$$T = \frac{2}{\pi} \int_0^\infty \int_0^\infty F(a, b, r, \lambda) f(\alpha) \sin(\lambda z) \sin(\lambda \alpha) d\lambda d\alpha. \quad (16)$$

satisfies all the conditions of the problem.

In the particular case where $f(z)$ reduces to a constant, T_0 , the solution (16) becomes

$$T = \frac{2}{\pi} T_0 \int_0^\infty F(a, b, r, \lambda) \sin(\lambda z) d\lambda \int_0^\infty \sin(\lambda \alpha) d\alpha. \quad (17)$$

Since the integrated expression of $\int_0^\infty \sin(\lambda \alpha) d\alpha$ is indeterminate at the upper limit, it is necessary to rewrite solution (17) in the form

$$\begin{aligned} T &= \lim_{\rho \rightarrow \infty} \frac{2}{\pi} T_0 \int_0^\infty F(a, b, r, \lambda) \sin(\lambda z) d\lambda \int_0^\rho \sin(\lambda \alpha) d\alpha \\ &= \lim_{\rho \rightarrow \infty} \frac{2}{\pi} T_0 \int_0^\infty F(a, b, r, \lambda) \sin(\lambda z) \frac{1 - \cos(\lambda \rho)}{\lambda} d\lambda \\ &= \frac{2}{\pi} T_0 \int_0^\infty F(a, b, r, \lambda) \frac{\sin(\lambda z)}{\lambda} d\lambda \\ &\quad - \lim_{\rho \rightarrow \infty} \frac{2}{\pi} T_0 \int_0^\infty F(a, b, r, \lambda) \frac{\sin(\lambda z) \cos(\lambda \rho)}{\lambda} d\lambda. \end{aligned}$$

We shall prove that as ρ tends to ∞ , the second term of the last expression vanishes. This expression can be written in the form

$$\frac{1}{\pi} T_0 \int_0^\infty F(a, b, r, \lambda) \frac{\sin \lambda(\rho+z)}{\lambda} d\lambda - \frac{1}{\pi} T_0 \int_0^\infty F(a, b, r, \lambda) \frac{\sin \lambda(\rho-z)}{\lambda} d\lambda.$$

By the well-known property of the "Dirichlet Integrals," each term of the last expression tends to $F(a, b, r, 0)$ as $\rho \rightarrow \infty$, and, therefore, the expression under consideration vanishes. The solution for $f(z) = T_0$ thus finally becomes

$$T = \frac{2}{\pi} T_0 \int_0^\infty \frac{I_0(\lambda r) K_0(\lambda a) - I_0(\lambda a) K_0(\lambda r)}{I_0(\lambda b) K_0(\lambda a) - I_0(\lambda a) K_0(\lambda r)} \frac{\sin(\lambda z)}{\lambda} d\lambda.$$

Yeshiva College,
New York City.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

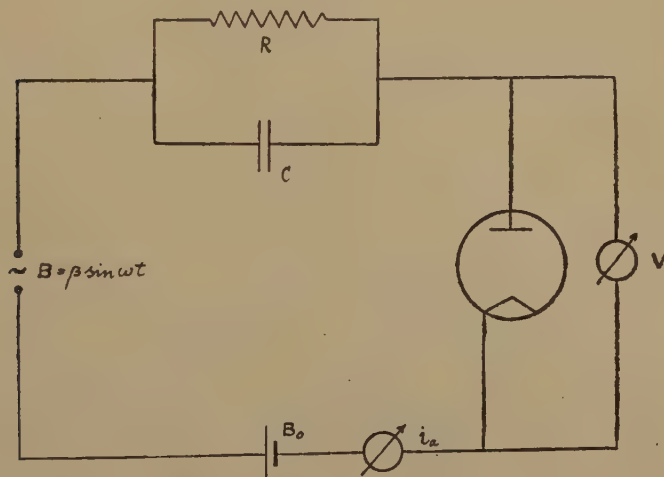
LIX. *The Diode as a Rectifier and a Detector for Weak Signals.*

By J. AHARONI, King's College, London, at H. H. Wills' Laboratory,
University of Bristol *.

[Received October 19, 1942.]

IN many problems connected with the radio valve it is permissible to limit the attention to the linear part of the characteristic. The linear circuit theory can then be applied. But in certain types of problem

Fig. 1.



the deviation of the characteristic from a straight line has to be taken into account. In such cases we meet with non-linear differential equations, and the linear circuit theory is of no direct help. In most cases these differential equations are of a complicated nature and methods of approximation have to be applied. It appears that the method of perturbation (applied so much in planetary and atomic problems) offers a convenient approach. The deviation from a straight line may be regarded as a perturbation, and a solution is sought in the form of a power series in the perturbation, and not in the applied signal voltage as in Carson's method ⁽¹⁾.

PART I.—RECTIFICATION.

Fig. 1 shows the diode in a typical circuit. Assume first the capacity $C=0$. We will also assume throughout that the characteristic of the valve is given by a quadratic function. This will be the case up to a

* Communicated by Dr. H. T. Flint.

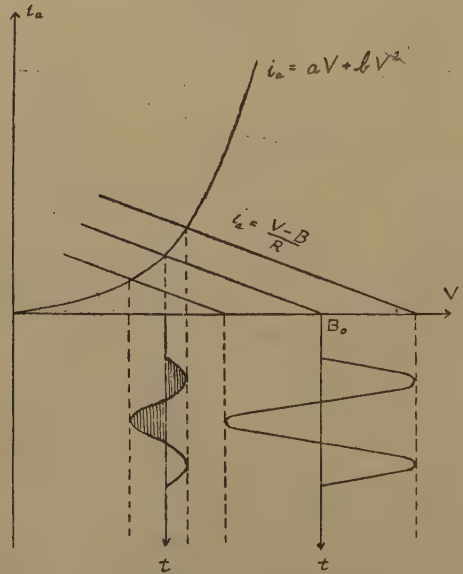
certain range of voltages. Then, if the frequencies involved are not too high, and the transit time of the electrons does not make itself felt, the static characteristic can be used, and the current through the valve and the voltage across it will be determined from the following equations:

$$i_a = aV + bV^2, \quad \dots \dots \dots (1)$$

$$V = B_0 + B - i_a R. \quad \dots \dots \dots (2)$$

a , b are constants which determine the characteristic, $B = \beta \sin \omega t$ is the applied signal, B_0 a constant bias. In this case ($C=0$) it is possible to give a graphical description of the problem by plotting the characteristic (1) and the straight line (2). Their intersection gives the solution (fig. 2).

Fig. 2.



As B is a periodic function of time, the straight line moves to and fro parallel to itself, the intercept on the V -axis being $B_0 + \beta \sin \omega t$. The extreme positions of the line are shown in fig. 2, and two auxiliary time axes, pointing downwards, are also shown. When $B=0$ the straight line goes through the point B_0 on the V -axis. At each moment the voltage V at the anode can be obtained by projecting the corresponding intersection on the V -axis. Because of the non-linearity of the characteristic the voltage at the anode will not be sinusoidal. As can be seen from the figure, the anode, which is originally positive by a value smaller than B_0 , becomes on the average less positive when the voltage $B = \beta \sin \omega t$ is applied owing to the shape of the characteristic. The shift on the average to a less positive value represents the rectifying effect of the diode.

Algebraically, if we solve equations (1) and (2), we get

$$V = -\frac{aR+1}{2bR} + \sqrt{\frac{(aR+1)^2}{4b^2R^2} + \frac{B_0+B}{bR}},$$

$$= \frac{aR+1}{2bR} \left[-1 + \sqrt{1 + \frac{(B_0+B)4bR}{(aR+1)^2}} \right] \dots \dots (3)$$

(The + sign of the square root must be taken because, for $B=0$, V must be positive.)

In the expression for V , B is equal to $\beta \sin \omega t$, and because of the square root, the voltage at the anode is not sinusoidal. V is of the form

$$V = p + \sqrt{q + r \sin \omega t},$$

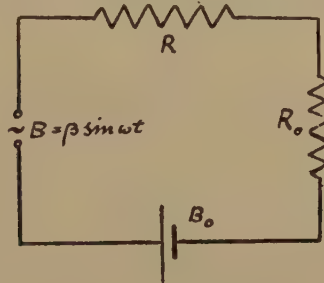
and the average \bar{V} ,

$$\bar{V} = \frac{1}{\tau} \int_0^\tau V dt, \text{ where } \tau = \frac{2\pi}{\omega},$$

will be

$$\bar{V} = p + \frac{1}{\tau} \int_0^\tau \sqrt{q + r \sin \omega t} dt.$$

Fig. 3.



If b , which measures the deviation from linearity, is small, or if by a suitable choice of R we can achieve

$$\frac{4bR(B+B_0)}{(aR+1)^2} < 1,$$

then we can expand the square root in (3),

$$V = \frac{B+B_0}{aR+1} - \frac{(B+B_0)^2 bR}{(aR+1)^3} + \dots \dots \dots (4)$$

The first term represents the voltage that would develop if the characteristic were linear. Then $a = \frac{1}{R_0}$ = the internal conductance of the diode and

$$\frac{B+B_0}{aR+1} = \frac{(B+B_0)R_0}{R+R_0}.$$

The equivalent circuit would be as shown in fig. 3.

The second term in (4) shows the effect of b . It gives a voltage term proportional to $(B+B_0)^2=(\beta \sin \omega t+B_0)^2$. The average will be

$$\frac{1}{\tau} \int_0^\tau (\beta \sin \omega t + B_0)^2 dt = \frac{1}{\tau} \int (\beta^2 \sin^2 \omega t + 2\beta B_0 \sin \omega t + B_0^2) dt = \frac{\beta^2}{2} + B_0^2$$

Hence the average voltage will be

$$\bar{V} = \frac{B_0}{aR+1} - \frac{bR B_0^2}{(aR+1)^3} - \frac{bR \beta^2}{2(aR+1)^3} \dots \dots \dots (5)$$

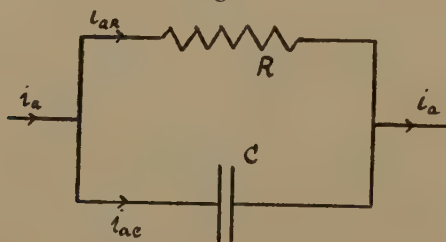
The first two terms give together the effect of the d.c. battery, taking into account the curvature of the characteristic to the first order. The last term gives the rectified voltage which will result when a sinusoidal E.M.F. is applied. It is proportional to the square of this E.M.F.

(4) may be rewritten as follows :

$$V = \frac{B_0}{aR+1} - \frac{bR \left(\frac{\beta^2}{2} + B_0^2 \right)}{(aR+1)^3} + \beta \left(\frac{1}{aR+1} - \frac{2B_0 bR}{(aR+1)^3} \right) \sin \omega t + \frac{bR \beta^2}{2(aR+1)^3} \cos 2\omega t. \quad (6)$$

Hence a second harmonic appears which is proportional to β^2 .

Fig. 4.



Next drop the assumption that the capacity of the condenser in fig. 1 is zero. Then we have to solve the following equations :

$$i_a = aV + bV^2,$$

$$i_a = i_{aC} + i_{aR},$$

$$\frac{1}{C} \int i_{aC} dt = i_{aR} R = B_0 + B - V.$$

$$\therefore \quad B - V = \frac{i_{aC}}{C} = \frac{i_a - i_{aR}}{C} = \frac{i_a}{C} - \frac{B_0 + B - V}{CR}.$$

$$\therefore \quad i_a = C(\dot{B} - \dot{V}) + \frac{B_0 + B - V}{R} = aV + bV^2,$$

$$\frac{dV}{dt} + \left(\frac{1}{CR} + \frac{a}{C} \right) V + \frac{bV^2}{C} = \frac{dB}{dt} + \frac{B + B_0}{CR},$$

$$\frac{dV}{dt} + \left(\frac{1}{CR} + \frac{a}{C} \right) V + \frac{bV^2}{C} = \alpha \cos \omega t + \bar{\beta} \sin \omega t + \gamma = \phi(t), \dots \dots (7)$$

where $\alpha = \beta\omega$, $\bar{\beta} = \frac{\beta}{CR}$, $\gamma = \frac{B_0}{CR}$.

Equation (7) can be written as follows :

$$\frac{dV}{dt} + mV + nV^2 = \phi(t), \quad (8)$$

where $m = \frac{1}{CR} + \frac{a}{C}$, $n = \frac{b}{C}$

Let us write the solution V in the form

$$V = V_0(t) + bV_1(t) = V_0(t) + f(t), \quad (9)$$

where $V_0(t)$ satisfies the differential equation (7) or (8) when $b=0$,

$$\frac{dV_0}{dt} + mV_0 = \phi(t). \quad (10)$$

Inserting (9) in (8) and using (10) we get

$$\frac{df}{dt} + mf + n(V_0 + f)^2 = 0. \quad (11)$$

We will now assume b to be sufficiently small for the terms in b^2 to be neglected. This means that terms in f^2 can be neglected or terms in bf (or $\frac{b}{C}f = nf$ no matter what C is).

Neglecting terms with f^2 (11) becomes

$$\frac{df}{dt} + (m + 2nV_0)f + nV_0^2 = 0. \quad (12)$$

This is a linear differential equation of the first order, of the type that would be met in a linear circuit with variable parameters.

Because of the linearity the term nV_0^2 (the force) can be written as a sum of terms in any manner we like and the differential equation solved for each term separately. The sum of the separate solutions is the solution of (12). It will be convenient to regard f as a current in an $L_1 R$ circuit ($L=1$ Henry, $R=a$ variable resistance $=m+2nV_0(t)$, $nV_0^2(t)$ being the applied force).

As regards V_0 the steady state solution of (10) is

$$\begin{aligned} V_0 &= e^{-\int m dt} \int \phi(t) e^{\int m dt} dt, \\ &= e^{-mt} \int (\alpha \cos \omega t + \bar{\beta} \sin \omega t + \gamma) e^{mt} dt. \end{aligned}$$

Now $\int \cos \omega t e^{mt} dt = \frac{m}{\omega^2 + m^2} e^{mt} \cos \omega t + \frac{\omega}{\omega^2 + m^2} e^{mt} \sin \omega t,$

so that V_0 becomes

$$V_0 = \frac{\alpha m - \bar{\beta} \omega}{\omega^2 + m^2} \cos \omega t + \frac{\alpha \omega + \bar{\beta} m}{\omega^2 + m^2} \sin \omega t + \frac{\gamma}{m},$$

say $P_1 \cos \omega t + P_2 \sin \omega t + \frac{\gamma}{m}$.

Let $P_1 = A \sin \psi, \quad P_2 = A \cos \psi,$

then $V_0 = A \sin (\omega t + \psi) + \frac{\gamma}{m}, \quad (13)$

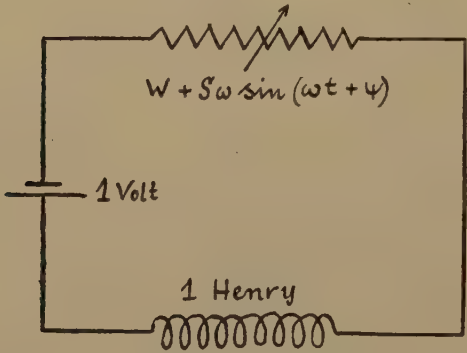
where $A = \sqrt{\left(\frac{\alpha m - \beta \omega}{\omega^2 + m^2}\right)^2 + \left(\frac{\alpha \omega + \beta m}{\omega^2 + m^2}\right)^2} \quad (14)$

and $\tan \psi = \frac{P_1}{P_2} = \frac{\alpha m - \beta \omega}{\alpha \omega + \beta m} \quad (15)$

The steady state solution of (12) will be given by

$$f = -e^{-\int (m + 2nV_0) dt} \int nV_0^2 e^{\int (m + 2nV_0) dt} dt. \quad (16)$$

Fig. 5.



Now, using (13),

$$m + 2nV_0 = \frac{2n\gamma + m^2}{m} + 2nA \sin (\omega t + \psi)$$

and $\int (m + 2nV_0) dt = \frac{2n\gamma + m^2}{m} t - \frac{2nA}{\omega} \cos (\omega t + \psi)$

$$= Wt - S \cos (\omega t + \psi),$$

where $W = \frac{2n\gamma + m^2}{m} \quad \text{and} \quad S = \frac{2nA}{\omega}.$

(16) now becomes

$$\begin{aligned} f &= -e^{-Wt + S \cos (\omega t + \psi)} \int nA^2 \sin^2 (\omega t + \psi) e^{Wt - S \cos (\omega t + \psi)} dt \\ &\quad - e^{-Wt + S \cos (\omega t + \psi)} \int 2nA \frac{\gamma}{m} \sin (\omega t + \psi) e^{Wt - S \cos (\omega t + \psi)} dt \\ &\quad - e^{-Wt + S \cos (\omega t + \psi)} \int \frac{n\gamma^2}{m^2} e^{Wt - S \cos (\omega t + \psi)} dt \\ &= I_1 + I_2 + I_3. \quad (17) \end{aligned}$$

Let us call

$$-e^{-(Wt+S \cos(\omega t+\psi))} \int e^{Wt-S \cos(\omega t+\psi)} dt = \Phi(t). \quad (18)$$

Numerically it is equal to the current which will flow in the steady state in the circuit shown in fig. 5, because $\Phi(t)$ is the steady solution of

$$\frac{d\Phi}{dt} + (W + S\omega \sin(\omega t + \psi))\Phi + 1 = 0.$$

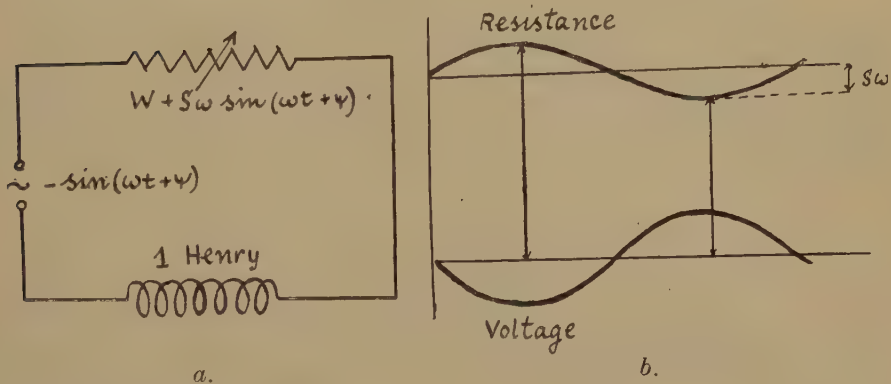
Using 18, I_3 becomes

$$I_3 = n \frac{\gamma^2}{m^2} \Phi(t). \quad (19)$$

Next consider I_2 . Writing

$$-e^{-Wt+S \cos(\omega t+\psi)} \int \sin(\omega t+\psi) e^{Wt-S \cos(\omega t+\psi)} dt = X(t), \quad (20)$$

Fig. 6.



we shall have

$$I_2 = 2nA \frac{\gamma}{m} X(t). \quad (21)$$

The function $X(t)$ is numerically equal to the current which will flow in the steady state in the circuit of fig. 6a.

Voltage and resistance fluctuation are in antiphase so that when the voltage is positive the resistance is low, and when the voltage is negative the resistance is high. The result is that in the steady state, apart from a periodic current component a d.c. component will also flow,

$$\frac{1}{\tau} \int_0^\tau \overline{X(t)} dt = X(t) \neq 0.$$

It remains to consider I_1

$$\begin{aligned} & -e^{-Wt+S \cos(\omega t+\psi)} \int \sin^2(\omega t+\psi) e^{Wt-S \cos(\omega t+\psi)} dt \\ &= -e^{-Wt+S \cos(\omega t+\psi)} \int \left[\frac{1}{2} - \frac{\cos(2\omega t+2\psi)}{2} \right] e^{Wt-S \cos(\omega t+\psi)} dt \\ &= \frac{1}{2} \Phi(t) - \frac{1}{2} \Psi(t), \end{aligned}$$

where

$$\Psi(t) = -e^{-Wt + S \cos(\omega t + \psi)} \int \cos(2\omega t + 2\psi) e^{Wt - S \cos(\omega t + \psi)} dt. \quad (22)$$

$$\therefore I_1 = \frac{nA^2}{2} \Phi(t) - \frac{nA^2}{2} \Psi(t). \quad (23)$$

$$\therefore f = \frac{nA^2}{2} \Phi(t) - \frac{nA^2}{2} \Psi(t) + 2nA \frac{\gamma}{m} X(t) + \frac{n\gamma^2}{m^2} \Phi(t). \quad (24)$$

As f is proportional to $n = \frac{b}{C}$, which is of the first order, there is no need to calculate $\Phi(t)$, $\Psi(t)$ and $X(t)$ to the first order in $S = \frac{2nA}{\omega} = \frac{2bA}{\omega C}$, since terms of this and of higher orders result in terms of higher order in the function f . Therefore we may use

$$\begin{aligned} e^{Wt - S \cos(\omega t + \psi)} &= e^{Wt} e^{-S \cos(\omega t + \psi)} \doteq e^{Wt}, \\ e^{-Wt + S \cos(\omega t + \psi)} &= e^{-Wt} e^{S \cos(\omega t + \psi)} \doteq e^{-Wt}. \end{aligned}$$

This will also be clear from the following. In the circuits for I_1 , I_2 and I_3 the driving forces are of order b and hence the currents are of order b . The resistance fluctuation is also of order b , and this will cause a current fluctuation of order b^2 and hence can be neglected. As a result of this we have

$$\Phi(t) = -e^{-Wt} \int e^{Wt} dt = -\frac{1}{W}. \quad (25)$$

$$X(t) = -e^{-Wt} \int \sin(\omega t + \psi) e^{Wt} dt. \quad (26)$$

$$\Psi(t) = -e^{-Wt} \int \cos(2\omega t + 2\psi) e^{Wt} dt. \quad (27)$$

Let us first consider the average value of f . From the obvious circuit interpretation of (25), (26) and (27) it follows

$$\overline{\Phi(t)} = -\frac{1}{W}, \quad \overline{X(t)} = 0, \quad \overline{\Psi(t)} = 0,$$

($\Psi(t)$ has a frequency 2ω and the average for $\tau = \frac{2\pi}{\omega}$ will be zero), and from (24)

$$\frac{1}{\tau} \int_0^\tau f dt = \bar{f} = -\frac{nA^2}{2W} - \frac{n\gamma^2}{m^2 W}. \quad (28)$$

$$\text{Now } W = \frac{2n\gamma + m^2}{m} = \frac{(1 + aR)^2 + 2bB_0R}{(1 + aR)CR}. \quad (29)$$

$$A^2 = \left(\frac{\alpha m - \bar{\beta} \omega}{\omega^2 + m^2} \right)^2 + \left(\frac{\alpha \omega + \bar{\beta} m}{\omega^2 + m^2} \right)^2 = \beta^2 \left[\frac{1 + \omega^2 R^2 C^2}{(1 + aR)^2 + \omega^2 R^2 C^2} \right]. \quad (30)$$

$$\frac{\gamma}{m} = \frac{B_0}{1 + aR}. \quad (31)$$

$$\frac{1}{W} = \frac{(1+aR)CR}{(1+aR)^2 + 2bB_0R} = \frac{CR}{(1+aR) \left(1 + \frac{2bB_0R}{(1+aR)^2}\right)} \quad (32)$$

$$\frac{n}{W} = \frac{b}{CW} = \frac{bRC}{C(1+aR) \left(1 + \frac{2bB_0R}{(1+aR)^2}\right)} = \frac{bR}{1+aR} \quad (33)$$

(to the first order in b).

$$\tan \psi = \frac{\alpha m - \bar{\beta} \omega}{\alpha \omega + \bar{\beta} m} = \frac{\omega a R^2 C}{\omega^2 R^2 C^2 + (1+aR)} \quad (34)$$

Eq. (28) becomes

$$\bar{f} = -\frac{\beta^2}{2} \frac{(1+\omega^2 R^2 C^2)}{(1+aR)^2 + \omega^2 R^2 C^2} \frac{bR}{1+aR} - \frac{B_0^2}{(1+aR)^2} \frac{bR}{1+aR} \quad (35)$$

The average of $V = V_0 + f$ will be

$$\bar{V} = \bar{V}_0 + \bar{f} = \frac{B_0}{1+aR} - \frac{B_0 b^2 R}{(1+aR)^3} - \frac{\beta^2}{2} \frac{(1+\omega^2 R^2 C^2)}{(1+aR)^2 + \omega^2 R^2 C^2} \frac{bR}{1+aR} \quad (36)$$

The last term will be equal for $C=0$ to that obtained before (see eq. (5)). (36) shows that the larger the capacity the bigger will be the rectified voltage. The maximum value will be

$$-\frac{\beta^2}{2} \frac{bR}{1+aR},$$

whereas without a condenser the rectified voltage is

$$-\frac{\beta^2}{2} \frac{bR}{(1+aR)^3},$$

The main advantage gained from the use of the condenser is its smoothing effect. To obtain this we must calculate $\Phi(t)$, $X(t)$ and $\Psi(t)$ from (25), (26) and (27).

$$\Phi(t) = -\frac{1}{W},$$

as we have already seen.

$$\begin{aligned} X(t) &= -e^{-Wt} \int \sin(\omega t + \psi) e^{Wt} dt \\ &= \left[\frac{-\omega}{\omega^2 + W^2} \cos(\omega t + \psi) + \frac{W}{\omega^2 + W^2} \sin(\omega t + \psi) \right] \\ X(t) &= -\frac{1}{\sqrt{4\omega^2 + W^2}} \sin(\omega t + \psi + \phi_1), \end{aligned} \quad (37)$$

where

$$\tan \phi_1 = -\frac{\omega}{W} \quad (38)$$

Similarly, from (27),

$$\Psi(t) = -\frac{1}{\sqrt{4\omega^2 + W^2}} \cos(2\omega t + 2\psi + \phi_2), \quad (39)$$

where $\tan \phi_2 = -\frac{2\omega}{W}$.

Inserting (25), (38), and (39) in (24) we get

$$f = -\frac{nA^2}{2W} + \frac{nA^2}{2} \frac{\cos(2\omega t + 2\psi + \phi_2)}{\sqrt{4\omega^2 + W^2}} - \frac{2nA\gamma}{m} \frac{1}{\sqrt{\omega^2 + W^2}} \sin(\omega t + \psi + \phi_1) - \frac{n\gamma^2}{m^2} \cdot \frac{1}{W} \quad (40)$$

In (40)

$$\frac{n}{\sqrt{\omega^2 + W^2}} = \frac{bR}{\sqrt{\omega^2 C^2 R^2 + (1 + aR)^2}} \quad \dots \quad (41a)$$

(to the first order in b), and

$$\frac{n}{\sqrt{4\omega^2 + W^2}} = \frac{bR}{\sqrt{4\omega^2 C^2 R^2 + (1 + aR)^2}} \quad \dots \quad (41b)$$

The amplitude F_1 of the fundamental frequency as contributed by f will be

$$F_1 = 2A \frac{\gamma}{m} \frac{n}{\sqrt{\omega^2 + W^2}} = \frac{2\beta R \sqrt{1 + \omega^2 R^2 C^2}}{\omega^2 C^2 R^2 + (1 + aR)^2} \frac{bB_0}{1 + aR} \quad \dots \quad (42)$$

The amplitude as contributed by V_0 will be

$$A = \beta \left[\frac{1 + \omega^2 R^2 C^2}{(1 + aR)^2 + \omega^2 R^2 C^2} \right]^{\frac{1}{2}}$$

Now we cannot simply add A and F_1 as V_0 and the relevant part in f are not in phase. But for $\omega RC \gg 1 + aR$, ϕ_1 , the difference in phase between the relevant part in f and V_0 , becomes $= 0$ (see (38) and (13)), so that the amplitude of the fundamental across the valve, F_v , is

$$F_v = \beta - \frac{2\beta}{\omega C} \frac{bB_0}{1 + aR} \quad (\omega RC \gg 1 + aR) \quad \dots \quad (43)$$

Had the characteristic been linear, the amplitude of the fundamental across the valve would have been β , and across the load ($B - V$) nearly zero. It is not quite zero because V_0 and $B = \beta \sin \omega t$ are not in phase. However, if C is large ($\omega RC \gg 1 + aR$) the phase difference ψ becomes

$$\frac{a}{\omega C} = \frac{aR}{\omega CR} \div 0, \text{ so that had the characteristic been linear perfect smoothing}$$

would have been achieved (with no rectification of course). From (43) it will be seen that the deviation of the characteristic from a straight line makes things slightly worse as regards smoothing proportionately with b . The rectified voltage will be

$$\frac{\beta^2}{2} \frac{bR}{1 + aR} \quad (\omega RC \gg 1 + aR)$$

and the amplitude of the fundamental frequency across the load will be

$$\frac{2\beta}{\omega C} \frac{bB_0}{1 + aR}.$$

The efficiency of the smoothing will be given by the percentage ripple

$$\frac{4B_0}{\beta\omega CR} 100 \text{ (percentage ripple).}$$

The amplitude of the harmonic 2ω will be

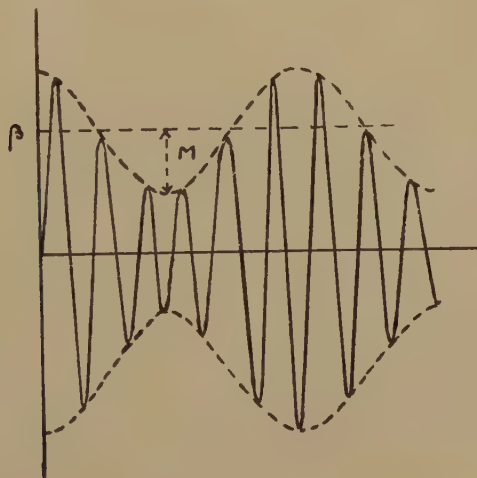
$$\frac{b\beta^2}{4\omega C}$$

and the percentage ripple will be

$$\frac{b(1+aR)}{4\omega CR} 100.$$

It may be added that for $C=0$, $\psi=0$, $\phi_1=0$, $\phi_2=0$ and V_0+f becomes the same expression as obtained before (eq. 6).

Fig. 7.



PART II.—DEMODULATION OR DETECTION.

Assume the voltage B applied to the diode to consist of a modulated signal (fig. 7),

$$B = (\beta + M \sin pt) \sin \omega t, \quad . \quad . \quad . \quad . \quad . \quad (44)$$

ω being the carrier frequency and p the audio-frequency. The expression (44) may also be written in the following form :

$$B = \beta \sin \omega t + \frac{M}{2} \cos (\omega - p)t - \frac{M}{2} \cos (\omega + p)t.$$

It contains three frequencies : ω , $\omega - p$ and $\omega + p$, the carrier and the two side-bands. These latter are 180° out of phase relative to one another.

Sometimes it is not sufficiently emphasized that the modulated wave does not contain a component of audio-frequency. The fact is that amplitude-modulated wave represents only the raw material out of which

From Part I we know that

$$V_{01} = A_1 \sin(\omega t + \psi_1) + \frac{\gamma}{m}, \quad . \quad . \quad . \quad . \quad . \quad (51)$$

$$V_{02} = A_2 \sin((\omega - p)t + \psi_2) + \frac{\gamma}{m}, \quad . \quad . \quad . \quad . \quad . \quad (52)$$

$$V_{03} = A_3 \sin((\omega + p)t + \psi_3) + \frac{\gamma}{m}. \quad . \quad . \quad . \quad . \quad (53)$$

So that

$$\begin{aligned} & V_{01}V_{02} + V_{02}V_{03} + V_{01}V_{03} \\ &= A_1A_2 \sin(\omega t + \psi_1) \sin((\omega - p)t + \psi_2) \\ & \quad + A_2A_3 \sin((\omega - p)t + \psi_2) \sin((\omega + p)t + \psi_3) \\ & \quad + A_1A_3 \sin(\omega t + \psi_1) \sin((\omega + p)t + \psi_3) \\ & \quad + \frac{2\gamma}{n} [A_1 \sin(\omega t + \psi_1) + A_2 \sin((\omega - p)t + \psi_2) \\ & \quad + A_3 \sin((\omega + p)t + \psi_3)] + \frac{3\gamma^2}{m^2}, \\ &= \frac{A_1A_2}{2} \cos(pt + \psi') - \frac{A_1A_2}{2} \cos((2\omega - p)t + \psi) \\ & \quad + \frac{A_2A_3}{2} \cos(2pt + \psi'') - \frac{A_1A_3}{2} \cos(2\omega t + \psi) \\ & \quad + \frac{A_1A_3}{2} \cos(pt + \psi''') - \frac{A_1A_3}{2} \cos((2\omega + p)t + \psi) \\ & \quad + \frac{2\gamma}{m} [A_1 \sin(\omega t + \psi_1) + A_2 \sin((\omega - p)t + \psi_2) \\ & \quad + A_3 \sin((\omega + p)t + \psi_3)] + \frac{3\gamma^2}{m^2}, \quad . \quad . \quad . \quad . \quad . \quad (54) \end{aligned}$$

where

$$\begin{aligned} \psi' &= \psi_1 - \psi_2, & \psi'' &= \psi_2 - \psi_3, & \psi''' &= \psi_1 - \psi_3, \\ \psi' &= \psi_1 + \psi_2, & \psi'' &= \psi_3 + \psi_2, & \psi''' &= \psi_1 + \psi_3. \end{aligned} \quad . \quad . \quad . \quad . \quad . \quad (55)$$

As the differential equation for f_p in (50) is linear, each term in the sum (54) will give rise to an additive part in the solution. For convenience, let us regard f_p as a current and the sum (54) as the applied E.M.F. As can be seen from (54) the applied force consists of ten separate sources of frequencies, $2\omega + p$, 2ω , $2\omega - p$, $\omega + p$, ω , $\omega - p$, $2p$, p and d.c. Only the sources with p and $2p$ are audio-frequencies, the rest are of radio-frequencies or d.c., and are therefore of no further interest at present. The production of the frequency p is what is wanted, whereas the oscillation with the frequency $2p$ is a by-product which is not desirable. The second harmonic of the lower audio-frequencies will still be audible, and its appearance will certainly modify the quality of the sound. It is important to consider the proportionality of the amplitude of the audio-frequency p to the depth of the modulation of the incident wave.

The terms with frequency p are

$$\begin{aligned} & \frac{A_1 A_2}{2} \cos (pt + \psi') + \frac{A_1 A_3}{2} \cos (pt + \psi''') \\ &= \left[\frac{A_1 A_2}{2} \cos \psi' + \frac{A_1 A_3}{2} \cos \psi''' \right] \cos pt \\ & \quad - \left[\frac{A_1 A_2}{2} \sin \psi' + \frac{A_1 A_3}{2} \sin \psi''' \right] \sin pt = L \sin (pt + \phi), \quad (56) \end{aligned}$$

where

$$\begin{aligned} L &= \sqrt{\left[\frac{A_1 A_2}{2} \cos \psi' + \frac{A_1 A_3}{2} \cos \psi''' \right]^2 + \left[\frac{A_1 A_2}{2} \sin \psi' + \frac{A_1 A_3}{2} \sin \psi''' \right]^2} \\ L &= \sqrt{\frac{A_1^2 A_2^2}{2} + \frac{A_1^2 A_3^2}{2} + \frac{1}{2} A_1^2 A_2 A_3 (\cos \psi' \cos \psi''' + \sin \psi' \sin \psi''')}, \quad (57) \end{aligned}$$

and because of $\psi' - \psi'' = \psi_3 - \psi_2$,

$$L = A_1 \sqrt{\frac{A_2^2}{2} + \frac{A_3^2}{2} + \frac{1}{2} A_2 A_3 \cos (\psi_2 - \psi_3)}. \quad (58)$$

A_1 results from the carrier frequency $\beta \sin \omega t = B_1$. Using (30),

$$A_1 = \beta \sqrt{\frac{1 + \omega^2 R^2 C^2}{(1 + aR)^2 + \omega^2 R^2 C^2}}.$$

We will now make the assumption that the condenser is such that

$$\omega RC \gg 1 + aR,$$

then

$$A_1 \doteq \beta. \quad (59)$$

A_2 results from the lower side-band

$$\begin{aligned} B_2 &= \frac{M}{2} \cos ((\omega - p)t) = \frac{M}{2} \sin \left((\omega - p)t + \frac{\pi}{2} \right) \\ A_2 &= \frac{M}{2} \sqrt{\frac{1 + (\omega - p)^2 R^2 C^2}{(1 + aR)^2 + (\omega - p)^2 R^2 C^2}} \end{aligned}$$

and, as $\omega - p \doteq \omega$,

$$A_2 \doteq \frac{M}{2}, \quad (60)$$

and in the same way

$$A_3 \doteq \frac{M}{2}. \quad (61)$$

In (58) the difference $\psi_2 - \psi_3$ appears. It is clear that as the capacity is large it will affect both side-bands in the same way, so that $\psi_2 - \psi_3$ will equal the initial phase difference between the two side-bands

$$\psi_2 - \psi_3 = \pi.$$

(This can be checked also by using (34). For large C where

$$\omega RC \gg 1 + aR, \quad \tan \psi \doteq \frac{aR}{\omega CR} \ll 1.$$

$$\text{Thus } \psi \doteq \frac{a}{\omega C}, \quad \psi_2 = \frac{a}{(\omega - p)C} + \frac{\pi}{2}, \quad \psi_3 = \frac{a}{(\omega + p)C} - \frac{\pi}{2}, \quad \psi_2 - \psi_3 \doteq \pi.$$

As $\psi_2 - \psi_3 = \pi$, (56) becomes

$$L = \beta \sqrt{\frac{M^2}{8} + \frac{M^2}{8} - \frac{M^2}{8}} = \frac{\beta M \sqrt{2}}{4}. \quad (62)$$

The audio-frequency part of f_p , let us call it f_d , will thus satisfy the following differential equation:

$$\frac{df_d}{dt} + [m + 2n(V_{01} + V_{02} + V_{03})]f_d + \frac{2n\beta M \sqrt{2}}{4} \sin(pt + \phi) = 0. \quad (63)$$

We know already that, as the applied force in (63) is proportional to n , and hence to b , we can ignore the fluctuations in the resistance term, and to the first order in b ,

$$\begin{aligned} f_d &= -e^{-mt} \int \frac{2n\beta M \sqrt{2}}{4} \sin(pt + \phi) e^{mt} dt, \\ f_d &= -\frac{2n\beta M \sqrt{2}}{4\sqrt{p^2 + m^2}} \sin(pt + \phi + \phi'), \\ f_d &= \frac{-\sqrt{2}b\beta MR}{2\sqrt{p^2 R^2 C^2 + (1 + aR)^2}} \sin(pt + \phi + \phi'). \quad (64) \end{aligned}$$

This formula shows that the amplitude of the audio-frequency oscillation, which has been produced by the diode, is proportional both to β , the amplitude of the carrier wave, and to M , the amplitude of the modulation. Hence, to the first order, the diode behaves as an ideal demodulator. The proportionality both with β and M makes it clear that it is better to amplify before detection than use the same amplification after detection. If we amplify the arriving wave with a radio frequency amplifier n times, both β and M will become n times as big, so that βM will be amplified by n^2 . But should we amplify the audio-frequency by n the result will remain an amplification by a factor n .

The appearance of p in the square root of (62), shows that for different audio-frequencies, corresponding all to the same modulation depth, the higher frequencies will be produced with a smaller amplitude. But as long as

$$pRC \ll 1 + aR,$$

this smoothing effect will not be appreciable.

From some text-books the impression is gained that as the rectification is according to a square law so will also be the detection. It is perhaps therefore worth while to show by an elementary consideration that, exactly because of the square law rectification, the demodulation is linear.

Assume the incident carrier wave to be modulated by a square wave (fig. 8)

In each part we shall have a rectification according to a square law, and the result will be as shown in fig. 9.

The amplitude of the square wave produced will be

$$D = \frac{1}{2}[(\beta + M)^2 - (\beta - M)^2] = 2\beta M,$$

so that the proportionality both with β and M is demonstrated. A linear demodulation would follow, of course, also from a linear rectification.

Finally, a few words should be said about the accuracy of the result. We have treated b as if it were in our means to make it as small as we like. Mathematically this is a convenient way of dealing with the differential equation (7). But it is clear that it is the relative value of b that matters. Equation (7) may be written as follows:—

$$C \frac{dV}{dt} + \left(a + \frac{1}{R}\right)V + bV^2 = \frac{B}{R} + \frac{B_0}{R} + C\dot{B},$$

Fig. 8.

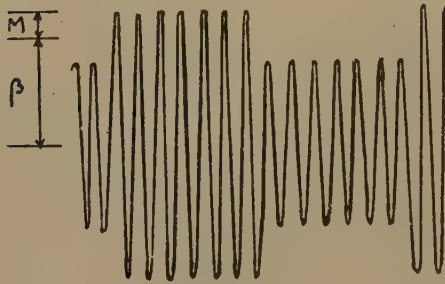


Fig. 9.



and it is clear that if our solution V , to the first order, is such that

$$\epsilon = \frac{bV^2R}{(1+aR)V} = \frac{bVR}{1+aR} \ll 1, \quad \dots \dots \dots (65)$$

the solution V to the first order will be accurate to a high degree and terms in b^2 really negligible. Then, using the first order result, we know that for, $\omega RC \gg 1+aR$, $|V| = |V_0| + |f| = \beta + f \doteq \beta$

$$\epsilon = \frac{b\beta R}{1+aR} \ll 1. \quad \dots \dots \dots (66)$$

This is the condition which must be satisfied if terms in b^2 are to be neglected, and from which the accuracy of the result can be found. (66) shows that the weaker the signal β , the more will the higher terms in b be negligible, and hence the more will the rectification be according to a square law, and the demodulator linear,

By a suitable choice of R , ϵ can be made as small as we like, but it must be taken into account that from (64) pRC must be made small as compared with $1+aR$.

For the amplitude of f_a we get from (64)

$$|f_a| = \frac{\sqrt{2}M}{2} \frac{b\beta R}{1+aR} = \frac{\sqrt{2}}{2} M\epsilon. \quad \dots \quad (64a)$$

The straightening effect of R must not be made too large. The more fidelity we expect from the rectifier the less efficient it will be. Lack of fidelity will be proportional to ϵ^2 , efficiency to ϵ .

Reference.

(1) "The Equivalent Circuit of the Vacuum Tube Modulator." J. R. Carson, Proc. I. R. E. vol. ix. p. 243 (1921).

LX. *Reflexion and Transmission by Absorbing Dielectrics of Electromagnetic Waves in Hollow Tubes.*

By L. PINCHERLE, King's College, London.
At the H. H. Wills Physics Laboratory, Royal Fort, Bristol *.

[Received October 6, 1942.]

1. Introduction.

WITHIN the last few years much interest has arisen in the propagation of electromagnetic waves in hollow metal tubes—first studied by Rayleigh⁽¹⁾—both from an experimental and a theoretical point of view. The two principal features of electromagnetic waves in tubes are, first that only waves of a wave-length inferior to the cross-sectional dimension of the tube can be propagated (and when the dielectric in the tube is non-absorbing there is a critical frequency above which the waves are propagated without attenuation, and below which they are very strongly absorbed); and second that there is either an electric or a magnetic intensity in the direction of propagation of the waves, *i. e.* along the tube. When there is an electric intensity in the direction of propagation, the wave is called an E-wave; and when there is a longitudinal magnetic intensity, the wave is designated an H-wave. As shown by Brillouin⁽²⁾, the longitudinal component arises from the fact that the waves in the tube can be considered as the resultant of four (two in some special cases) transverse plane waves, travelling across the tube with the velocity corresponding to that of plane waves in an unrestricted medium, and reflected at the walls of the tube, thus constituting a number of criss-cross paths. These criss-cross waves will be denoted in the following as the "component waves."

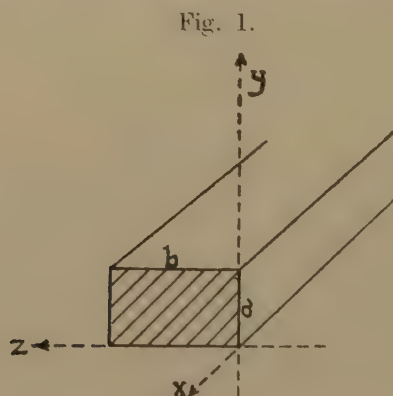
* Communicated by Dr. H. T. Flint.

The purpose of this note is to investigate theoretically the reflexion and transmission of the waves at the surface of separation, normal to the axis of the tube, between two different media. We shall consider tubes of rectangular cross-section, and suppose that the walls are infinitely conducting. It will be shown that the formulæ obtained for the reflected and transmitted intensities are identical with the usual Fresnel's formulæ, for the absorbing case, applied to the component waves.

In Section 2 the expressions for the fields in the absorbing dielectric inside a tube of rectangular cross-section will be given for convenience: they are the natural extension of those valid for the non-absorbing case. In Section 3 the boundary conditions at the separation of two media will be applied to find the reflected and transmitted intensities. In Section 4 formulæ relative to a sheet of an absorbing dielectric will be given, as they may be suitable for experimental applications.

2. The Fields in the Absorbing Dielectric.

We shall use the M.K.S. system of rationalized units. The system of co-ordinates and the dimensions of the tube are shown in fig. 1.



The system of co-ordinates and dimensions of the tube.

Starting from Maxwell's equations :

$$\left. \begin{aligned} \text{curl } \bar{H} &= \sigma \bar{E} + \epsilon \frac{\partial \bar{E}}{\partial t}; & \text{curl } \bar{E} &= -\mu \frac{\partial \bar{H}}{\partial t}; \\ \text{div } \bar{H} &= 0; & \text{div } \bar{E} &= 0; \end{aligned} \right\} \quad \dots \quad (1)$$

we assume the dependance of every component of the fields on t and x to be given by the expression :

$$e^{i\omega t - \alpha x}, \quad \dots \quad (2)$$

with $\omega = 2\pi f$ and $h = \alpha + i\beta$, where α is the attenuation and β the propagation constant ($\beta = \frac{2\pi}{\lambda}$ in vacuum).

Eliminating H_y , H_z (or E_y , E_z) from the first two rotational equations,

and substituting the values of E_y , E_z (or H_y , H_z) in the divergence equation, we get for E_x (or H_x) the equation :

$$\frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} + \gamma^2 E_x = 0, \quad \dots \dots \dots (3)$$

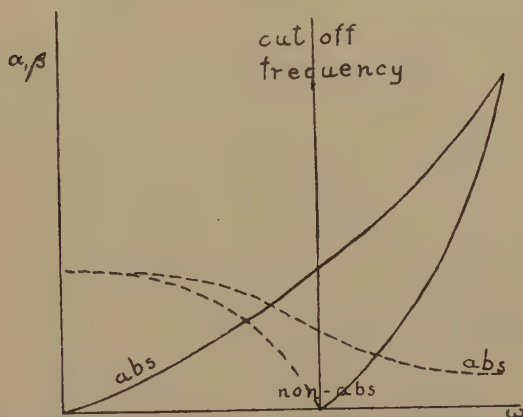
with $\gamma^2 = h^2 + \epsilon\mu\omega^2 - i\mu\sigma\omega. \quad \dots \dots \dots (4)$

In view of the linear character of equation (3), separate solutions can be obtained for E_x (with $H_x=0$) and H_x (with $E_x=0$), corresponding to E-waves and H-waves respectively ; a general wave will result from superimposition of individual waves.

The boundary conditions at the walls are : annulment of the tangential component of E , and annulment of the normal derivative of H . These conditions lead, both for E- and for H-waves, to the relation :

$$\gamma^2 = \left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2 \quad n, m=0, 1, 2, \dots \dots \dots (5)$$

Fig. 2.



The real (α , broken curve) and imaginary (β , continuous curve) part of h plotted against the frequency for an absorbing and a non-absorbing dielectric with the same dielectric constant.

To each couple of values n, m , corresponds an $E_{n,m}$ wave and an $H_{n,m}$ wave ; for an E-wave both m and n must be $\neq 0$; for an H-wave only $H_{0,0}$ is excluded. In all cases we have, from (4), (5) :

$$h = \sqrt{\gamma^2 - \epsilon\mu\omega^2 + i\sigma\mu\omega}, \quad \dots \dots \dots (6)$$

so that the attenuation and propagation constants are given by :

$$\left. \begin{aligned} \alpha &= \sqrt{\frac{1}{2}[\gamma^2 - \epsilon\mu\omega^2 + \sqrt{(\epsilon\mu\omega^2 - \gamma^2)^2 + \sigma^2\mu^2\omega^2}]} \\ \beta &= \sqrt{\frac{1}{2}[\epsilon\mu\omega^2 - \gamma^2 + \sqrt{(\epsilon\mu\omega^2 - \gamma^2)^2 + \sigma^2\mu^2\omega^2}]} \\ 2\alpha\beta &= \sigma\mu\omega \quad (3). \end{aligned} \right\} \quad \dots \dots (7)$$

For no absorption these reduce to the usual formulæ :

$$\alpha = 0 ; \quad \beta = \sqrt{\epsilon\mu\omega^2 - \gamma^2} \quad (\text{for } \epsilon\mu\omega^2 > \gamma^2), \quad \dots \dots (8)$$

and (7) approaches (8) as ω increases.

The importance of γ lies in the fact that it includes both the dimensions a , b of the tube and the order m , n of the wave, so that we have only one quantity instead of four.

In the absorbing case there is not a cut-off frequency for which the propagation constant vanishes as in the non-absorbing case, but the real part of h increases steadily for decreasing ω , until, for $\epsilon\mu\omega^2 \ll \gamma^2$, its value approaches the value for the non-absorbing case (fig. 2).

If $\sigma \ll \epsilon\omega$, (7) can be written

$$\beta = \sqrt{\epsilon\mu\omega^2 - \gamma^2}; \quad \alpha = \frac{\mu\sigma\omega}{2\sqrt{\epsilon\mu\omega^2 - \gamma^2}}; \quad . \quad . \quad . \quad (9)$$

if, moreover, ω is very large, or γ very small, then :

$$\beta = \sqrt{\epsilon\mu} \omega; \quad \alpha = \frac{\sigma}{2} \sqrt{\frac{\mu}{\epsilon}}.$$

The expressions for the fields are :—

E-waves :

$$\left. \begin{aligned} E_x &= A \sin\left(\frac{n\pi}{a}y\right) \sin\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \\ H_x &= 0; \\ E_y &= -\frac{h}{\gamma^2} A \frac{n\pi}{a} \cos\left(\frac{n\pi}{a}y\right) \sin\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \\ H_y &= \frac{Ai}{\gamma^2} (\omega\epsilon - i\sigma) \frac{m\pi}{b} \sin\left(\frac{n\pi}{a}y\right) \cos\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \\ E_z &= -\frac{h}{\gamma^2} A \frac{m\pi}{b} \sin\left(\frac{n\pi}{a}y\right) \cos\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \\ H_z &= -\frac{Ai}{\gamma^2} (\omega\epsilon - i\sigma) \frac{n\pi}{a} \cos\left(\frac{n\pi}{a}y\right) \sin\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \end{aligned} \right\} \quad (10)$$

H-waves :

$$\left. \begin{aligned} E_x &= 0; \\ H_x &= B \cos\left(\frac{n\pi}{a}y\right) \cos\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \\ E_y &= \frac{i\mu\omega}{\gamma^2} B \frac{m\pi}{b} \cos\left(\frac{n\pi}{a}y\right) \sin\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \\ H_y &= \frac{h}{\gamma^2} B \frac{n\pi}{a} \sin\left(\frac{n\pi}{a}y\right) \cos\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \\ E_z &= -\frac{i\mu\omega}{\gamma^2} B \frac{n\pi}{a} \sin\left(\frac{n\pi}{a}y\right) \cos\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}; \\ H_z &= \frac{h}{\gamma^2} B \frac{m\pi}{b} \cos\left(\frac{n\pi}{a}y\right) \sin\left(\frac{m\pi}{b}z\right) e^{-hx+i\omega t}. \end{aligned} \right\} \quad (11)$$

In both cases the vectors \bar{E} and \bar{H} are perpendicular.

The flux of energy through the pipe is given by the real part of the flux of the complex Poynting's vector, namely, for E-waves :

$$S_x = \text{Re} \int_0^a \int_0^b \frac{1}{2} (\mathbf{E}_y \mathbf{H}_z^* - \mathbf{E}_z \mathbf{H}_y^*) dy dz = \frac{ab A^2 e^{-2\alpha x}}{8\gamma^2} (\beta\omega\epsilon + \alpha\sigma), \quad (12)$$

and for H-waves :

$$S_x = \frac{ab B^2 e^{-2\alpha x}}{8\gamma^2} \beta\omega\mu. \quad (13)$$

The attenuation in decibel/meter is then given by :

$$d = 8.686\alpha. \quad (14)$$

For the case $\sigma \ll \epsilon\omega$, this takes the form

$$d = 4.343 \frac{\sigma\mu\omega}{\sqrt{\epsilon\mu\omega^2 - \gamma^2}}, \quad (15)$$

identical with formula (20) of Kemp's ⁽⁴⁾.

Component Waves.

If, in formulæ (10), (11), we express the functions such as $\sin\left(\frac{n\pi}{a}y\right)$ by means of exponentials, we obtain each field component as the sum of four parts (two in the special case of the $H_{0,m}$ or $H_{n,0}$ waves). These correspond to four criss-cross component waves, and the vector of propagation of these has the components

$$-\beta \quad \pm \frac{n\pi}{a} \quad \pm \frac{m\pi}{b},$$

so that in all four cases the propagation vector makes with the x -axis an angle ϑ such that

$$\sin \vartheta = \frac{\gamma}{\sqrt{\beta^2 + \gamma^2}}; \quad \cos \vartheta = \frac{\beta}{\sqrt{\beta^2 + \gamma^2}}. \quad (16)$$

The wave-length is, for all four,

$$\lambda = \frac{2\pi}{\sqrt{\beta^2 + \gamma^2}}. \quad (17)$$

These component waves are inhomogeneous ⁽⁵⁾, *i. e.* the planes of equal phase, normal to the vector of propagation, make an angle $\vartheta \neq 0$ with the planes of equal amplitude, normal to the x -axis. They propagate in the absorbing dielectric exactly in the same way as in an unbounded dielectric with the same constants. In fact, from (17), the refractive index of a given medium for these waves is given by

$$n = \frac{\lambda_0}{\lambda} = \frac{2\pi c}{\omega} \frac{\sqrt{\beta^2 + \gamma^2}}{2\pi} = \frac{c}{\omega} \sqrt{\beta^2 + \gamma^2}, \quad (18)$$

and the absorption index (notation as in König) :

$$k = \frac{\lambda}{2\pi} q'' = \frac{\lambda}{2\pi} \alpha \cos \vartheta = \frac{\alpha\beta}{\beta^2 + \gamma^2} = \frac{\sigma\mu\omega}{2(\beta^2 + \gamma^2)}. \quad (19)$$

Now for an inhomogeneous wave we have generally

$$n^2 = \frac{\epsilon\mu c^2}{2} \left[1 + \sqrt{1 + \left(\frac{\sigma}{\epsilon\omega \cos \vartheta} \right)^2} \right]; \quad \dots \quad (18')$$

$$k = \frac{\sigma}{\epsilon\omega} \left[1 + \sqrt{1 + \left(\frac{\sigma}{\epsilon\omega \cos \vartheta} \right)^2} \right]^{-1}, \quad \dots \quad (19')$$

and, remembering (7) and (16), it is easily seen that (18) and (18'), (19) and (19') coincide.

For no absorption we have $\beta^2 + \gamma^2 = \epsilon\mu\omega^2$, and the ordinary formulæ

$$\sin \vartheta = \frac{\gamma}{\sqrt{\epsilon\mu} \omega}; \quad \lambda = \frac{2\pi}{\sqrt{\epsilon\mu} \omega}; \quad n = c\sqrt{\epsilon\mu}^{(2)}.$$

When the dimensions of the tube increase ($\gamma \rightarrow 0$), we have

$$\sin \vartheta \simeq \frac{\gamma}{\beta}; \quad n \simeq \frac{c\beta}{\omega}; \quad k \simeq \frac{\sigma\mu\omega}{2\beta^2} = \frac{\alpha}{\beta}.$$

3. Reflexion and Transmission.

We suppose the tube to be filled with a dielectric of constants ϵ_1, σ_1 , for $x < 0$, and with a dielectric of constants ϵ_2, σ_2 , for $x > 0$. We take $\mu_2 = \mu_1 = \mu$ vacuum. The conditions for the continuity of the tangential components of \vec{E} and \vec{H} , and of the normal components of \vec{B} and \vec{D} at the surface of separation, applied to (10) and (11), give for the reflected amplitude:—

$$\text{E-waves: } \frac{A'}{A} = \frac{h_1\eta_2 - h_2\eta_1}{h_1\eta_2 + h_2\eta_1}, \quad \dots \quad (20)$$

$$\text{H-waves: } \frac{B'}{B} = \frac{h_1 - h_2}{h_1 + h_2}, \quad \dots \quad (21)$$

where $\eta = \omega\epsilon - i\sigma$,

and h_1, h_2 are the (complex) propagation constants for the two media.

Similarly, for the refracted wave*:—

$$\text{E-waves: } \frac{A''}{A} = \frac{2h_1\eta_1}{h_1\eta_2 + h_2\eta_1}, \quad \dots \quad (22)$$

$$\text{H-waves: } \frac{B''}{B} = \frac{2h_1}{h_1 + h_2}. \quad \dots \quad (23)$$

These reflexion and transmission coefficients are obtained also by the use of the concept of impedance, as introduced by Schelkunoff⁽⁶⁾.

Let us consider the case in which the first dielectric is non-absorbing ($\sigma_1 = 0$). Then the expressions for the reflected and refracted intensities are:

$$\text{E-waves: } r_e^2 = \left| \frac{A'}{A} \right|^2 = \frac{(\sigma_2\beta_1 - \omega\epsilon_1\alpha_2)^2 + \omega^2(\epsilon_2\beta_1 - \epsilon_1\beta_2)^2}{(\sigma_2\beta_1 + \omega\epsilon_1\alpha_2)^2 + \omega^2(\epsilon_2\beta_1 + \epsilon_1\beta_2)^2}, \quad \dots \quad (24)$$

* We use the expression "refraction" besides "transmission," as the component waves suffer an ordinary refraction, as is apparent from (16) and (18).

$$\text{H-waves: } r_h^2 = \left| \frac{B'}{B} \right|^2 = \frac{\alpha_2^2 + (\beta_1 - \beta_2)^2}{\alpha_2^2 + (\beta_1 + \beta_2)^2}, \quad \dots \quad (25)$$

while there is a change of phase ϕ at the reflexion given by

$$\text{E-waves: } tg\phi_e = \frac{2\epsilon_1\omega\beta_1(\omega\epsilon_2\alpha_2 - \sigma_2\beta_2)}{\beta_1^2\sigma_2^2 + \alpha_2^2\epsilon_1^2\omega^2 + \omega^2(\epsilon_2^2\beta_1^2 - \epsilon_1^2\beta_2^2)}, \quad \dots \quad (26)$$

$$\text{H-waves: } tg\phi_h = -\frac{2\alpha_2\beta_1}{\alpha_2^2 - \beta_1^2 + \beta_2^2}. \quad \dots \quad (27)$$

The reflected intensity can never vanish for H-waves, and it can be proved that, until σ_2 is $\neq 0$, it cannot vanish for E-waves either, as the two conditions

$$\epsilon_2\beta_1 = \epsilon_1\beta_2; \quad \sigma_2\beta_1 = \omega\epsilon_1\alpha_2, \quad \dots \quad (28)$$

are never satisfied together.

When the dimensions of the tube increase ($\gamma \rightarrow 0$), both (24) and (25) reduce to the ordinary formula for normal incidence in an unrestricted medium.

The conservation of energy is expressed by:

$$\text{E-waves: } \left| \frac{A'}{A} \right|^2 + \frac{\beta_2}{\beta_1} \left[\frac{\epsilon_2}{\epsilon_1} + \frac{\sigma_2\alpha_2}{\epsilon_1\omega\beta_2} \right] \left| \frac{A''}{A} \right|^2 = 1, \quad \dots \quad (29)$$

$$\text{H-waves: } \left| \frac{B'}{B} \right|^2 + \frac{\beta_2}{\beta_1} \left| \frac{B''}{B} \right|^2 = 1. \quad \dots \quad (30)$$

The proof that the preceding formulæ are identical with the ordinary formulæ for reflexion and refraction by an unrestricted absorbing dielectric, when these ordinary formulæ are applied to the component waves, is immediate. We suppose that the first medium is air $\left(\epsilon_1\mu_1 = \frac{1}{c^2} \right)$.

To H-waves corresponds an electric vector perpendicular to the plane of incidence, so that the reflected intensity is given by ⁽⁵⁾:

$$\left| \frac{B'}{B} \right|^2 = \frac{\sin^2(i-r) + \frac{k^2 \sin^2 i}{\cos^2 r}}{\sin^2(i+r) + \frac{k^2 \sin^2 i}{\cos^2 r}}, \quad \dots \quad (31)$$

where all quantities refer to the component waves.

Remembering (16) and (19), we have

$$\sin i = \frac{\gamma c}{\omega}, \quad \sin r = \frac{\gamma}{\sqrt{\beta_2^2 + \gamma^2}}, \quad k^2 = \frac{\alpha_2\beta_2}{\beta_2^2 + \gamma^2},$$

and remembering that for air $\beta_1 = \sqrt{\frac{\omega^2}{c^2} - \gamma^2}$, (31) becomes

$$\left| \frac{B'}{B} \right|^2 = \frac{(\beta_2 - \beta_1)^2 + \alpha_2^2}{(\beta_2 + \beta_1)^2 + \alpha_2^2},$$

identical with (25).

Similarly, it is easily seen that the change of phase at reflexion

$$tg\phi_h = -k \frac{2 \sin i \sin r \cos i \cos r}{\sin(i+r) \sin(i-r) \cos^2 r + k^2 \sin^2 i}$$

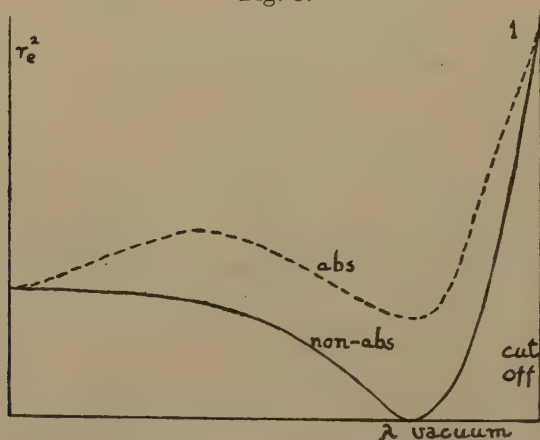
is identical with (27).

For E-waves the proof requires a somewhat longer calculation, as the formulæ for this case are complicated, but the result is the same.

As (24) is simpler than the usual formulæ, it can be used also for the reflexion by an unbounded dielectric: it is necessary to enclose the dielectric in a fictitious tube of such dimensions that, if waves were propagated along it, the component waves would have the required angle of incidence i ; to obtain this we must take $\gamma = \frac{\omega}{c} \sin i$, and then calculate $\beta_1, \alpha_2, \beta_2$ by means of (7).

We have now seen that the consideration of the component waves enables us to apply the ordinary laws also for reflexion and refraction at an absorbing dielectric; so that the behaviour of waves in tubes, although somewhat strange, can always be reduced to well-known phenomena.

Fig. 3.



Intensity of the E-wave reflected by a non-absorbing (continuous curve) and by an absorbing (broken curve) dielectric, with the same dielectric constant plotted against the wave-length in vacuum.

Case of no Absorption.

In this case all formulæ simplify and no complex quantities are involved. For instance, we have for the reflected amplitude, instead of (19) and (20):

$$\text{E-waves: } \frac{A'}{A} = \frac{\beta_1 \epsilon_2 - \beta_2 \epsilon_1}{\beta_1 \epsilon_2 + \beta_2 \epsilon_1}, \quad \dots \dots \dots (32)$$

$$\text{H-waves: } \frac{B'}{B} = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}. \quad \dots \dots \dots (33)$$

These are easily shown to be the same as Fresnel's formulæ for the component waves. In particular, no reflexion occurs for E-waves when $\epsilon_1 \beta_2 = \epsilon_2 \beta_1$, corresponding to Brewster's angle of incidence for the component waves. Introducing the cut-off frequency for the first medium

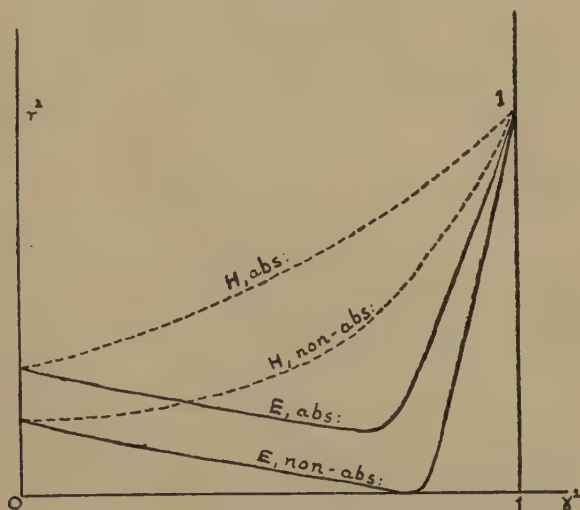
$$f_1 = \frac{\gamma}{2\pi \sqrt{\epsilon_1 \mu_1}}, \quad \dots \dots \dots (34)$$

and supposing, as usual, $\mu_2 = \mu_1$, we find that no reflexion occurs when the frequency of the waves is

$$f_B = f_1 \sqrt{1 + \frac{\epsilon_1}{\epsilon_2}}. \quad \dots \dots \dots (35)$$

Fig. 3 gives the intensity of the reflected wave, plotted against the vacuum wave-length for a non-absorbing, and for an absorbing substance, both for an E-wave. Fig. 4 gives the same plotted against γ^2 for E- and H-waves.

Fig. 4.



Reflected intensity by a dielectric of infinite thickness for an E- and an H-wave as a function of γ^2 . Values of the constants: $\omega = 3 \cdot 10^8$, $\epsilon_1 = \epsilon_0$, $\epsilon_2 = 5\epsilon_0$, $\mu_1 = \mu_2 = \mu_0$, $\sigma_2 = 10^{-2}$, $-\sigma_1 = 0$.

4. Dielectric Sheet.

We consider a dielectric sheet limited by plane surfaces normal to the axis of the tube at $x=0$ and $x=x_0$. The treatment is then the same as for an infinite plane sheet in an unrestricted medium.

We shall consider only the case of E-waves. The formulæ for H-waves are identical when we substitute B' , B'' , r_h , ϕ_h , for the corresponding quantities.

At the first surface the reflected amplitude is given by (20):

$$\frac{A'}{A} = \frac{h_1 \eta_2 - h_2 \eta_1}{h_1 \eta_2 + h_2 \eta_1} = r_e e^{i\phi_e}, \quad \dots \dots \dots (36)$$

where r_e and ϕ_e are given by (24) and (26).

The transmitted amplitude is given by (22):

$$\frac{A''}{A} = \frac{2h_1 \eta_2}{h_1 \eta_2 + h_2 \eta_1} = t_e e^{i\psi_e}, \quad \dots \dots \dots (37)$$

with

$$t_e = \frac{2\omega \epsilon_1 \beta_1}{\sqrt{(\sigma_2 \beta_1 + \omega \epsilon_1 \alpha_2)^2 + \omega^2 (\epsilon_1 \beta_2 + \epsilon_2 \beta_1)^2}}, \quad \dots \dots \dots (38)$$

$$\operatorname{tg} \psi_e = \frac{\omega \epsilon_1 \alpha_2 + \sigma_2 \beta_1}{\omega (\epsilon_1 \beta_2 + \epsilon_2 \beta_1)}. \quad (39)$$

At the second surface the corresponding formulæ are :

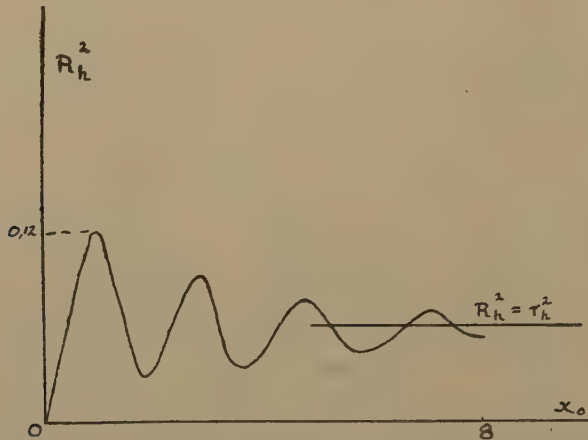
$$\frac{A'}{A} = -r_e e^{i\psi_e}; \quad \frac{A''}{A} = t_e' e^{i\psi_e'}, \quad (40)$$

with
$$t_e' = \sqrt{\frac{4(\omega \epsilon_2 \alpha_2 + \sigma_2 \beta_2)^2 + 4(\omega \epsilon_2 \beta_2 - \sigma_2 \alpha_2)^2}{(\omega \epsilon_1 \alpha_2 + \sigma_2 \beta_1)^2 + \omega^2 (\epsilon_2 \beta_1 + \epsilon_1 \beta_2)^2}}, \quad (41)$$

$$\operatorname{tg} \psi_e' = -\frac{\omega \epsilon_1 \sigma_2 (\alpha_2^2 + \beta_2^2) + (\epsilon_2^2 \omega^2 + \sigma_2^2) \alpha_2 \beta_1}{\omega^2 \epsilon_1 \epsilon_2 (\alpha_2^2 + \beta_2^2) + (\epsilon_2^2 \omega^2 + \sigma_2^2) \beta_1 \beta_2}, \quad (42)$$

and the relation holds
$$t_e t_e' e^{i(\psi_e + \psi_e')} + r_e^2 e^{2i\psi_e} = 1. \quad (43)$$

Fig. 5.



Values of the constants: $\omega = 3 \cdot 10^8$, $\gamma^2 = 0.64$, $\epsilon_1 = \epsilon_0$, $\epsilon_2 = 3\epsilon_0$, $\sigma_2 = 10^{-3}$.
Reflected intensity by a dielectric sheet of thickness x_0 for an H-wave.

The amplitude reflected by the dielectric sheet is then :

$$\begin{aligned} \frac{A'}{A} &= r_e e^{i\psi_e} \left[1 - t_e t_e' e^{i(\psi_e + \psi_e')} - 2h_2 x_0 \sum_{n=0}^{\infty} r_e^{2n} e^{2n(i\phi_e - h_2 x_0)} \right] \\ &= \frac{r_e e^{i\psi_e} (1 - e^{2h_2 x_0})}{r_e^2 e^{2i\psi_e} - e^{2h_2 x_0}} = R_e e^{i\psi_e}, \quad (44) \end{aligned}$$

with
$$R_e^2 = \frac{r_e^2 (1 + e^{4\alpha_2 x_0}) - 2r_e^2 e^{2\alpha_2 x_0} \cos 2\beta_2 x_0}{r_e^4 + e^{4\alpha_2 x_0} - 2r_e^2 e^{2\alpha_2 x_0} \cos (2\phi_e - 2\beta_2 x_0)}, \quad (45)$$

$$\operatorname{tg} \Phi_e = \frac{(r_e e^{4\alpha_2 x_0} - r_e^3) \sin \phi_e + (r_e^3 - r_e) e^{2\alpha_2 x_0} \sin (\phi_e - 2\beta_2 x_0)}{(r_e^3 + r_e e^{4\alpha_2 x_0}) \cos \phi_e - (r_e^3 + r_e) e^{2\alpha_2 x_0} \cos (\phi_e - 2\beta_2 x_0)}. \quad (46)$$

The transmitted amplitude is

$$\frac{A''}{A} = e^{-\alpha_2 x_0} \frac{1 - r_e^2 e^{2i\psi_e}}{1 - r_e^2 e^{2i\phi_e - 2h_2 x_0}} = T_e e^{i\psi_e}, \quad (47)$$

with
$$T_e^2 = e^{-2\alpha_2 x_0} \frac{1 + r_e^4 - 2r_e^2 \cos 2\phi_e}{1 + r_e^4 e^{-4\alpha_2 x_0} - 2r_e^2 e^{-2\alpha_2 x_0} \cos (2\phi_e - 2\beta_2 x_0)}, \quad (48)$$

$$tg\Psi_e = \frac{r_e^2 \sin(2\phi_e - 2\beta_2 x_0) - r_e^2 e^{2\alpha_2 x_0} \sin 2\phi_e + r_e^4 \sin 2\beta_2 x_0}{e^{2\alpha_2 x_0} - r_e^2 \cos(2\phi_e - 2\beta_2 x_0) - r_e^2 e^{2\alpha_2 x_0} \cos 2\phi_e + r_e^4 \cos 2\beta_2 x_0}. \quad (49)$$

Fig. 5 gives R_e^2 and fig. 6 gives T_e^2 plotted against the thickness x_0 of the sheet. There is no value of x_0 for which $R_e^2=0$.

Case of no Absorption.

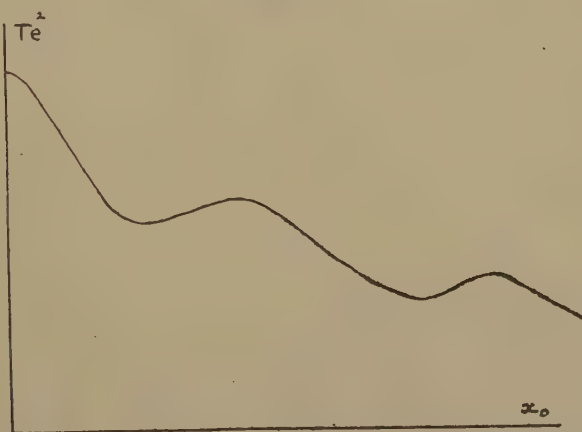
For no absorption the preceding formulæ become :

$$R_e^2 = \frac{2r_e^2(1 - \cos 2\beta_2 x_0)}{1 + r_e^4 - 2r_e^2 \cos 2\beta_2 x_0}; \quad (50)$$

$$T_e^2 = \frac{(1 - r_e^2)^2}{1 + r_e^4 - 2r_e^2 \cos 2\beta_2 x_0}, \quad (51)$$

with $R_e^2 + T_e^2 = 1$.

Fig. 6.



Transmitted intensity by a dielectric sheet of thickness x_0 .

Formulæ equivalent to these have been derived and verified experimentally by Lamont⁽⁷⁾ for a tube of circular cross-section.

When the thickness of the sheet varies, T_e^2 oscillates between the value 1 (for $\cos 2\beta_2 x_0 = 1$) and the value $\left(\frac{1 - r_e^2}{1 + r_e^2}\right)^2$ (for $\cos 2\beta_2 x_0 = -1$). Similarly R_e^2 oscillates between the values 0 and $(4r_e^2 / (1 + r_e^2)^2)$.

Always for no absorption :

$$tg\Phi_e = \frac{1 - r_e^2}{1 + r_e^2} \cotg \beta_2 x_0; \quad (52)$$

$$tg\Psi_e = -\frac{r_e^2 \sin 2\beta_2 x_0}{1 - r_e^2 \cos 2\beta_2 x_0}. \quad (53)$$

When βx_0 increases by $\frac{\pi}{2}$, also Φ_e increases by $\frac{\pi}{2}$, but in the interval the relation between the two angles is not linear.

All these expressions may be useful for the experimental determination of refractive and absorption indices by means of experiments in hollow tubes. Particularly formulæ (45), (46), can be applied when we investigate the position and amplitude of the maxima of the system of standing waves produced by reflexion on a sheet of variable thickness; or one of the methods introduced by Lamont can be used.

In conclusion, I wish to express my thanks to Dr. H. T. Flint for many helpful discussions.

Summary.

Formulæ for the reflected and transmitted intensities at the surface of separation between air and an absorbing dielectric, and by a sheet of this dielectric, are derived for electromagnetic waves in a hollow tube of rectangular cross-section with infinitely conducting walls.

It is shown that the same expressions are obtained by applying the ordinary Fresnel formulæ to the criss-cross waves into which the waves in the tube can be decomposed.

References.

- (1) Lord Rayleigh, *Phil. Mag.* xliii. p. 125 (1897).
G. C. Southwork, *Bell, Syst. Techn. Journ.* xv. p. 284 (1936).
J. R. Carson, S. P. Mead, and S. A. Schelkunoff, *Bell Syst. Techn. Journ.* xv. p. 310 (1936).
W. L. Barrow, *Proc. Inst. Radio Eng.* xxiv. p. 1298 (1936).
L. Brillouin, *Revue Gén. de l'Electricité*, xl. p. 227 (1936).
S. A. Schelkunoff, *Proc. Inst. Radio Eng.* xxv. p. 1457 (1937).
L. Page and N. I. Adams, *Phys. Rev.* lii. p. 647 (1937).
L. J. Chu and W. L. Barrow, *Proc. Inst. Radio Eng.* xxvi. p. 1520 (1938).
L. Brillouin, *Electrical Communication*, xvi. p. 350 (1938).
A. G. Clavier, *Electrical Communication*, xvii. p. 276 (1939).
H. R. L. Lamont, *Phil. Mag.* xxix. p. 521 (1939); xxx. p. 1 (1940).
J. Kemp, *Electr. Comm.* xx. p. 73 (1941).
V. C. A. Ferraro and H. T. Flint, *Proc. Phys. Soc.* liii. p. 170 (1941).
Other references in H. R. L. Lamont, 'Wave Guides,' Methuen and Co. (1942).
- (2) L. Brillouin, *Revue Gén. de l'Electricité*, xl. p. 227 (1936).
- (3) S. A. Schelkunoff, *Proc. Inst. Radio Eng.* xxv. pp. 1457-1468 (1937).
- (4) J. Kemp, *Electrical Communication*, xx. p. 73 (1941).
- (5) See *Handbuch der Physik*, Bd. xx. Article by W. König, p. 194; *l. c.* p. 204.
- (6) S. A. Schelkunoff, *Bell Syst. Techn. Journ.* xvii. p. 17 (1938).
- (7) H. R. L. Lamont, *Phil. Mag.* xxix. p. 521 (1940); xxx. p. 1 (1940).

LXI. *Inversion Formulæ for the Laplace Transformation.*

By A. ERDÉLYI, Mathematical Institute of the University,
Edinburgh *.

[Received February 8, 1943.]

1. ONE of the major problems in the practical applications of the operational calculus is the interpretation of operational images. The procedure in solving a problem by operational methods is generally as follows:—At first, by an application of the Laplace transformation, the problem is “translated” into a simpler problem in which, instead of the original functions, their operational images appear. This simpler problem is then solved and thereby the operational image of the solution of the original problem obtained. The final step consists in re-translating this operational image and thereby finding the solution of the original problem. It is this final step for which a method of interpreting operational images is necessary, *i. e.*, a method for inverting the Laplace transformation

$$g(s) = \mathfrak{L}[f(t)] \equiv \int_0^{\infty} e^{-st} f(t) dt, \quad (1)$$

or, yet in other words, a method for solving Laplace’s integral equation (1).

In many simple cases, for instance in electric circuit theory, a “dictionary” of Laplace transforms such as is contained in almost every book on operational calculus will suffice. If the operational image is either contained in such a “dictionary,” or else can be derived from functions appearing there, then the corresponding “object-function” is readily found. In other, more involved cases, especially in applications of the Laplace transformation to the solution of partial differential equations, there is no adequate “dictionary”: in such cases an *inversion formula* is needed, a general rule that is to say for solving (1), *i. e.*, finding $f(t)$ given $g(s)$. Accordingly there has grown up a whole literature dealing with the solution of (1). Most of the relevant publications are written from the point of view of pure mathematics and need not be discussed here; yet we must at least mention some of the more important inversion formulæ in order to compare them with those which form the main object of this note.

The best known and simplest of all formulæ for solving (1) is the so-called complex inversion formula: it is often associated with the name Mellin though it was substantially established by Cauchy and known to Riemann. This formula is so simple and general that it is used almost to the exclusion of the others. Its only drawback from the

* Communicated by the Author.

point of view of practical applications is that it involves complex integration and requires the knowledge of $g(s)$ along a line parallel to the imaginary axis. Inversion formulæ in form of an integral over part of the *real* axis have been found by Paley and Wiener⁽⁵⁾ (see also Doetsch⁽³⁾), Titchmarsh⁽⁶⁾, and, recently, by Boas and Widder⁽²⁾. These formulæ too, while useful in some cases, present difficulties in others: their application for numerical calculations is made difficult by the fact that they involve either a double integration or, in the last-mentioned case, a limiting process performed on an infinite integral. Other inversion formulæ determine $f(t)$ in form of an infinite series, not integral, the coefficients involved being obtained from the derivatives of $g(s)$. The practical application of such inversion formulæ is naturally restricted to those cases in which a general formula for the n th derivative of $g(s)$ is conveniently obtained. As Bateman⁽¹⁾ pointed out, such formulæ are of little use when $g(s)$ is found by graphical or numerical methods. A fairly general method is to expand either $f(t)$ or $g(s)$ in a series of orthogonal functions: several inversion formulæ can be obtained in this way (see⁽¹⁾ and the literature quoted there); the inversion formulæ proposed in the present note belong to this class. The particular feature of the present inversion formulæ is that the coefficients of the expansion in orthogonal functions of $f(t)$ are finite linear combinations of the values assumed by $g(s)$ at a set of equidistant points. Thus it is not necessary to know the values of $g(s)$ along the whole of the positive real axis; for the application of our inversion formulæ only a discrete set of values of $g(s)$ is relevant. Therefore these formulæ might be expected to be useful when $g(s)$ is found by numerical methods.

2. Let us put $e^{-it} = \sin \frac{1}{2}\theta$ in (1) and expand

$$\phi(\theta) = \cot \frac{1}{2}\theta \sin^{2a} \frac{1}{2}\theta f(t) \quad \dots \quad (2)$$

in a Fourier cosine series,

$$\phi(\theta) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos n\theta, \quad \text{where } a_n = \frac{2}{\pi} \int_0^{\pi} \phi(\theta) \cos n\theta d\theta.$$

Now⁽⁴⁾,

$$\cos n\theta = F(-n, n; \tfrac{1}{2}; \sin^2 \tfrac{1}{2}\theta) = \sum_{m=0}^n \frac{(-n)_m (n)_m}{(\tfrac{1}{2})_m m!} \sin^{2m} \tfrac{1}{2}\theta,$$

where $(c)_0 = 1$, $(c)_m = c(c+1) \dots (c+m-1)$ if $m=1, 2, \dots$:

$$\text{hence} \quad a_n = \frac{2}{\pi} \sum_{m=0}^n \frac{(-n)_m (n)_m}{(\tfrac{1}{2})_m m!} \int_0^{\pi} \phi(\theta) \sin^{2m} \tfrac{1}{2}\theta d\theta,$$

or, reintroducing $f(t)$ and t ,

$$a_n = \frac{2}{\pi} \sum_{m=0}^n \frac{(-n)_m (n)_m}{(\tfrac{1}{2})_m m!} \int_0^{\infty} e^{-(a+m)t} f(t) dt,$$

$$\text{i. e.,} \quad a_n = \frac{2}{\pi} \sum_{m=0}^n \frac{(-n)_m (n)_m}{(\tfrac{1}{2})_m m!} g(a+m) \quad (n=0, 1, 2, \dots). \quad \dots \quad (3)$$

Hence the following inversion formula for (1) :

$$f(t)=e^{at} \tan \frac{1}{2}\theta \{ \frac{1}{2}a_0+a_1 \cos \theta +a_2 \cos 2\theta + \dots \}, \quad . \quad . \quad . \quad (4)$$

where a is a suitably chosen constant, $\sin \frac{1}{2}\theta=e^{-\frac{1}{2}t}$, and the coefficients a_n are given by (3). Clearly a_n depends only on the values $g(s)$ takes for $s=a+m$ ($m=0, 1, \dots, n$), and a_n is a linear combination of the first $n+1$ of these values.

In a similar way other inversion formulæ, *e. g.*,

$$f(t)=e^{(a-\frac{1}{2})t} \{ a_0 \cos \frac{1}{2}\theta + a_1 \cos \frac{3}{2}\theta + \dots \},$$

$$f(t)=e^{at} \{ a_1 \sin \theta + a_2 \sin 2\theta + \dots \}$$

and
$$f(t)=e^{(a-\frac{1}{2})t} \tan \frac{1}{2}\theta \{ a_0 \sin \frac{1}{2}\theta + a_1 \sin \frac{3}{2}\theta + \dots \}$$

may be obtained in which the coefficients are determined by relations similar to (3); in each case a_n depends on $g(a+m)$ ($m \leq n$). We shall not investigate here the last three inversion formulæ since their properties are similar to those of (4).

3. In order to obtain another inversion formula for (1), let us put $1-2e^{-t}=x$ and expand

$$\psi(x)=e^{(1-b)t}f(t) \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

in a series of Legendre polynomials. We obtain

$$\psi(x)=\sum_{n=0}^{\infty} b_n P_n(x), \text{ where } b_n=(n+\frac{1}{2}) \int_{-1}^1 \psi(x) P_n(x) dx.$$

Now ⁽⁷⁾,

$$P_n(x)=F\left(-n, n+1; 1; \frac{1-x}{2}\right)=\sum_{m=0}^n \frac{(-n)_m (n+1)_m}{(m!)^2} e^{-mt},$$

and hence

$$b_n=2(n+\frac{1}{2}) \sum_{m=0}^n \frac{(-n)_m (n+1)_m}{(m!)^2} \int_0^{\infty} e^{-(b+m)t} f(t) dt,$$

or

$$b_n=(2n+1) \sum_{m=0}^n \frac{(-n)_m (n+1)_m}{(m!)^2} g(b+m) \quad (n=0, 1, 2, \dots). \quad (6)$$

Hence we obtain our second inversion formula for (1),

$$f(t)=e^{(b-1)t} \{ b_0 P_0(x) + b_1 P_1(x) + b_2 P_2(x) + \dots \}, \quad . \quad . \quad . \quad (7)$$

where b is a constant, suitably chosen, $x=1-2e^{-t}$ and b_n is given by (6). Again, b_n depends only on $g(b), g(b+1), \dots, g(b+n)$.

4. A rigorous mathematical theory of these inversion formulæ is beyond the scope of the present note: it will be given elsewhere. It is sufficient here to mention that theorems on the convergence and summability of the series (4) and (7) can be deduced from the theory of Fourier series and Legendre series respectively.

A few generalizations of our inversion formulæ may at least be indicated. It is known that $\cos n\theta=T_n(\cos \theta)$, where T_n is the Tchebichef

polynomial. Now, both Legendre polynomials and Tchebichef polynomials are particular cases of the so-called ultra-spherical polynomials, which in their turn are a sub-class of Jacobi polynomials. It is therefore probable that the latter can be used in order to derive more general inversion formulæ for the Laplace transformation. This conjecture can easily be verified: it is even true that any system of orthogonal polynomials for a finite interval can similarly be used. All these formulæ, if more general than those given above and possibly of advantage in some particular cases, do not seem to offer any essentially new point of interest.

A further generalization of the results presented here is thought to be more important and would seem to deserve closer investigation. All the orthogonal polynomials originate when the sequence $1, x, x^2, \dots, x^n, \dots$ is orthogonalized with respect to a suitable weight-function. Let us now consider the sequence $e^{-\lambda_0 t}, e^{-\lambda_1 t}, e^{-\lambda_2 t}, \dots$. If $w(t)$ is a suitable weight-function, we can orthogonalize our sequence with respect to $w(t)$ in $(0, \infty)$ and thereby gain an orthogonal system. The coefficients in the expansion of $f(t)$ in terms of these orthogonal functions will be linear combinations of the $g(\lambda_m)$, and thus inversion formulæ will be obtained similar to (4) and (7) yet more general inasmuch as the points λ_m at which the values of $g(s)$ are required need not be equidistant: they may be real or complex.

5. There is another interesting consequence of our series for $f(t)$. Substituting (4) into (1) we have

$$g(s) = \sum'_{n=0}^{\infty} a_n \int_0^{\infty} e^{-(s-a)t} \tan \frac{1}{2} \theta \cdot \cos n\theta dt,$$

where the prime indicates that in summing only half of the term for $n=0$ is to be taken. Now, the integral in the last equation is

$$\begin{aligned} \int_0^{\pi} \sin^{2s-2a} \frac{1}{2} \theta \cos n\theta d\theta &= 2(-)^n \int_0^{\frac{1}{2}\pi} \cos^{2s-2a} u \cos 2nu du \\ &= \frac{(-)^n 2^{2a-2s} \pi \Gamma(2s-2a+1)}{\Gamma(s-a+n+1) \Gamma(s-a-n+1)}, \end{aligned}$$

and hence

$$g(s) = \sum'_{n=0}^{\infty} a_n \frac{(-)^n 2^{2a-2s} \pi \Gamma(2s-2a+1)}{\Gamma(s-a+n+1) \Gamma(s-a-n+1)} \cdot \dots \quad (8)$$

Since the coefficients a_n depend on $g(a+n)$, this is an *interpolation formula* for Laplace transforms resembling somewhat Whittaker's interpolation formula by means of the cardinal series. Corresponding interpolation formulæ follow from the other inversion formulæ given or mentioned in the preceding sections.

References.

- (1) H. Bateman, *Duke Math. Journ.* ii. pp. 569-577 (1936).
- (2) R. P. Boas, Jr., and D. V. Widder, *Duke Math. Journ.* vi. pp. 1-26 (1940).

- (3) G. Doetsch, *Math. Zeits.*, xlii, pp. 263-286 (1937).
- (4) C. F. Gauss, *Gesammelte Werke*, iii. p. 137 (xx.).
- (5) R. E. A. C. Paley and N. Wiener, "Fourier Transformations in the Complex Domain" (*Am. Math. Soc. Coll. Publ.* vol. xix.), 1934.
- (6) E. C. Titchmarsh, 'Theory of Fourier Integrals,' Oxford, 1937, p. 316.
- (7) E. T. Whittaker and G. N. Watson, 'Modern Analysis,' Cambridge, 1920, p. 312.

LXII. On the Capture of a Slow-moving Directed Electron in a Coulomb Field of Force.

By W. R. MORGANS, D.Sc.*

[Received November 2, 1942.]

IN the treatment of problems involving the capture of electrons it is of advantage to use parabolical co-ordinates ⁽¹⁾. If we use the parabolical co-ordinates

$$x = \sqrt{\xi\eta} \cos \phi, \quad y = \sqrt{\xi\eta} \sin \phi, \quad z = \frac{\xi - \eta}{2},$$

$$0 \leq \xi \leq \infty, \quad 0 \leq \eta \leq \infty, \quad 0 \leq \phi \leq 2\pi,$$

and assume the electron and the nucleus as one system, then the Schrödinger wave equation for the electron travelling in the direction of the nuclear charge Ze is given by

$$\frac{\partial}{\partial \xi} \left(\xi \frac{\partial \psi}{\partial \xi} \right) + \frac{\partial}{\partial \eta} \left(\eta \frac{\partial \psi}{\partial \eta} \right) + \frac{1}{4} \left(\frac{1}{\xi} + \frac{1}{\eta} \right) \frac{\partial^2 \psi}{\partial \phi^2} + \frac{2\pi^2 \mu E}{h^2} (\xi + \eta) \psi + \frac{4\pi^2 \mu e^2 Z}{h^2} \psi = 0,$$

where ψ is the proper function, μ and e the mass and charge of the electron, E the energy of the atomic system and h is Plank's constant.

The Proper Functions.

In order to obtain the wave function corresponding to the electron moving in the direction of the z -axis with energy $E > 0$, Temple ⁽²⁾ solved the above wave equation by omitting the angular term and obtained a solution corresponding to the initial state of the electron in which the wave associated has symmetry about the z -axis. If we make the substitutions

$$k^2 = \frac{8\pi^2 \mu E}{h^2}, \quad a = \frac{h^2}{4\pi^2 \mu e^2 Z}, \quad n = \frac{1}{ak},$$

the solution found for the initial state ⁽³⁾ becomes

$$\psi(n) = A(n) e^{\frac{ik}{2}(\xi - \eta)} F(in, 1, ik\eta),$$

* Communicated by the Author.

where n is the parameter characterizing the continuous state, $A(n)$ the normalization factor

$$|A(n)|^2 = \frac{2\pi\mu a}{h} \cdot e^{\pi n} n \Gamma(1-in) \Gamma(1+in),$$

and where

$$F(\alpha, j, x) = \sum_{p=0}^{\infty} \frac{\alpha(\alpha+1) \dots (\alpha+p-1)}{j(j+1) \dots (j+p-1)} \frac{x^p}{p!}.$$

The wave function for the final state of the electron, in which the energy $E_1 < 0$, is given by the general solution of the preceding wave equation. If we make the substitution

$$k'^2 = -\frac{8\pi^2\mu E_1}{h^2},$$

the solution ⁽⁴⁾ is

$$\psi(n_1 n_2 m) = A(n_1 n_2 m) (\xi \eta)^{\frac{|m|}{2}} \epsilon^{-\frac{k'}{2}(\xi+\eta)}.$$

$$F(-n_1, |m|+1, k'\xi) F(-n_2, |m|+1, k'\eta) \frac{\cos}{\sin} m\phi,$$

where $|m| = 0, 1, 2, \dots$

$$k'(|m|+1+n_1+n_2) = \frac{1}{a} = k'l,$$

and $A(n_1 n_2 m)$, the normalization factor for the final state, is given by

$$[A(n_1 n_2 m)]^2 = \frac{2a}{\pi} \cdot \frac{(k')^{2|m|+4}}{\Gamma(|m|+1)^4} \frac{\Gamma(|m|+n_1+1) \Gamma(|m|+n_2+1)}{\Gamma(n_1+1) \Gamma(n_2+1)}.$$

The Matrices.

The components of the Heisenberg matrices are given by

$$x(n; n_1 n_2 m) = \int x \psi(n_1 n_2 m) \psi^*(n) dt,$$

with similar expressions for $y(n; n_1 n_2 m)$, $z(n; n_1 n_2 m)$, and where $\psi^*(n)$ denotes the conjugate value of $\psi(n)$. These matrices have been evaluated ⁽⁵⁾. If we use the relations $kn = k'l = \frac{1}{a}$, the matrices may be written as

$$x(n; n_1 n_2 1) = \hat{A}(n) (-1)^{n_1+n_2} \sqrt{\frac{2a}{\pi} (n_1+1)(n_2+1)} \frac{2\pi a^2 l^2 n^4}{(n^2+l^2)^2} (in) \\ v^{n_1-n_2-in} (1-v^2) F(1-in, -n_2; 2, 1-v^2),$$

where

$$v = \frac{k'-ik}{k'+ik} = \frac{n-il}{n+il},$$

F is the hypergeometric function, and where this matrix only remains finite for

$$x(n; n_1 n_2 m) = x(n; n_1 n_2 1) = x(n; n_1 n_2 -1).$$

The same expression holds for

$$y(n; n_1 n_2 m) = y(n; n_1 n_2 1) = y(n; n_1 n_2 -1).$$

The z -matrix is given by

$$z(n; n_1 n_2 0) = A(n) (-1)^{n_1+n_2} \sqrt{\frac{2a}{\pi}} \frac{2^2 \pi a^2 l^2 n^4}{(n^2 + l^2)^2} v^{n_1-n_2-in}$$

$$[\{l+v(n_1+1)\}F(-in, -n_2; 1, 1-v^2) + n_2 v F(-in, -n_2+1; 1, 1-v^2)],$$

where the matrix only remains finite for

$$z(n; n_1 n_2 m) = z(n; n_1 n_2 0).$$

Probability of Spontaneous Transition and the Absorption Coefficients.

The actual intensities associated with transitions between the states (n) and $(n_1 n_2 m)$ are proportional to the squares of the matrix elements x, y, z . However, when neglecting the relativity and the spin effects the energy of the unperturbed atom depends only upon n or

$$l = |m| + 1 + n_1 + n_2,$$

so that in order to obtain the intensity of transition for a given energy change, the squared matrix element must be summed over all the $n_1 n_2$. Owing to the relations

$$x(n; n_1 n_2 1) = x(n; n_1 n_2 - 1) = y(n; n_1 n_2 1) = y(n; n_1 n_2 - 1),$$

the intensity of the radiation may be written as

$$\begin{aligned} I &= \frac{64\pi^4 e^2}{3c^3} \nu^4 \left\{ 4 \sum_{n_2=0}^{l-2} |x(n; l-n_2-2, n_2, 1)|^2 \right. \\ &\quad \left. + \sum_{n_2=0}^{l-1} |z(n; l-n_2-1, n_2, 0)|^2 \right\} \\ &= \frac{64\pi^4 e^2}{3c^3} \nu^4 |J|^2, \end{aligned}$$

where ν , the frequency, is given by

$$\nu = \frac{E(n) - E_1(l)}{h} = \frac{h}{8\pi^2 \mu a^2} \frac{(n^2 + l^2)}{n^2 l^2}.$$

The probability of spontaneous transition A_{nl} per atom per unit time is given by

$$A_{nl} = \frac{I}{h\nu} = \frac{64\pi^4 e^2}{3hc^3} \nu^3 |J|^2,$$

and the coefficient of absorption α_ν per atom by

$$\alpha_\nu = \frac{8\pi^3 e^2}{3hc} \nu |J|^2.$$

The Lyman state for the atom for which $l=1$ gives $x(n; n_1 n_2 1)=0$ and $z(n; n_1 n_2 0)=z(n; 000)$. If we make use of the relation

$$\nu = e^{2i \arctan \frac{l}{n}},$$

we find

$$A_{n1} = \frac{2^6 \pi h e^2}{3c^3 \mu^2} \frac{n^6}{(n^2+1)^2} \frac{e^{-4n \arctan \frac{1}{n}}}{(1-e^{-2\pi n})}$$

and

$$\alpha_v = \frac{2\hbar^7}{3\pi^4\mu^4e^6Z^4c} \frac{n^{10}}{(n^2+1)^4} \frac{\epsilon^{-4n \arctan \frac{1}{n}}}{(1-\epsilon^{-2\pi n})}.$$

The Balmer state for the atom for which $l=2$, gives the following possibilities :

$$\begin{aligned} x(n; n_1n_21) & \quad n_1=n_2=0, \\ z(n; n_1n_20) & \quad n_1=1 \quad n_2=0, \\ & \quad n_1=0 \quad n_2=1. \end{aligned}$$

The squared matrices are found to be

$$\begin{aligned} |x(n; 001)|^2 &= \frac{2^{15}\pi^3\mu a^6}{\hbar} \frac{n^{14}}{(n^2+4)^6} \frac{\epsilon^{-4n \arctan \frac{2}{n}}}{(1-\epsilon^{-2\pi n})}, \\ |z(n; 100)|^2 &= \frac{2^{15}\pi^3\mu a^6}{\hbar} \frac{n^{12}}{(n^2+4)^5} \frac{\epsilon^{-4n \arctan \frac{2}{n}}}{(1-\epsilon^{-2\pi n})}, \\ |z(n; 010)|^2 &= \frac{2^{15}\pi^3\mu a^6}{\hbar} \frac{n^{12}(5n^2+4)^2}{(n^2+4)^7} \frac{\epsilon^{-4n \arctan \frac{2}{n}}}{(1-\epsilon^{-2\pi n})}, \end{aligned}$$

and the probability of spontaneous transition becomes

$$A_{n2} = \frac{2^7\pi e^2\hbar}{3\mu^2c^3} \frac{n^6(5n^2+4)(3n^2+4)}{(n^2+4)^4} \frac{\epsilon^{-4n \arctan \frac{2}{n}}}{(1-\epsilon^{-2\pi n})},$$

with

$$\alpha_v = \frac{2^6\hbar^7}{3\pi^4ce^6\mu^4Z^4} \frac{n^{10}(5n^2+4)(3n^2+4)}{(n^2+4)^6} \frac{\epsilon^{-4n \arctan \frac{2}{n}}}{(1-\epsilon^{-2\pi n})}.$$

The expressions for the matrices are convenient for numerical computation. If l is small the hypergeometric functions reduce to a few terms. If l is large the major contribution to the intensity comes from the square of the z matrix which contains the greatest numerical value of n_2 . Even when $l=3$, corresponding to the Paschen state, the contribution of $|z(n; 020)|^2$ is 81 per cent. of the total intensity. When l is large the method of steepest descent enables us to evaluate the matrices and the *complete* intensities for slow-moving electrons where $n \rightarrow \infty$.

Limiting Values.

(a) We shall consider the intensity for large l and for slow-moving electrons, so that we have to evaluate the limit of the expression

$$\begin{aligned} [l+v(n_1+1)]F(-in, -n_2; 1, z) \\ \sim lF(-in, -n_2; 1, z) \quad z=1-v^2 \end{aligned}$$

for $n \rightarrow \infty$, $n_2 \gg 1$.

$$\text{Now} \quad \lim F = \lim \frac{(-1)^{n_2}}{2\pi i} \int_c t^{-n_2-1} (1-t)^{n_2} (1-tz)^{in} dt,$$

where the path is a circuit around the pole $t=0$.

$$v = \frac{n-il}{n+il}; \quad z = 1 - v^2 = \frac{4inl}{(n+il)^2}$$

$$\lim z = \frac{4il}{n} \lim (1-zt)^n = \epsilon^{4it}.$$

Hence
$$\begin{aligned} \lim F &= \frac{(-1)^{n_1}}{2\pi i} \int_c t^{-n_1-1} (1-t)^{n_1} \epsilon^{4it} dt \\ &= \frac{(-1)^{n_1}}{2\pi i} \int_c t^{n_1-1} (1-t)^{l-n_1-1} e^{4it} dt. \end{aligned}$$

Put
$$\begin{aligned} e^\phi &= t^{-1} (1-t) e^{4it}, \\ \phi &= -\log t + \log (1-t) + 4it \end{aligned}$$

and find the saddle point $\frac{d\phi}{dt} = 0$.

$$\frac{d\phi}{dt} = -\frac{1}{t} + \frac{1}{t-1} + 4i, \quad \frac{d\phi}{dt} = 0 \text{ for } t=t_0=\frac{1}{2},$$

$$\frac{d^2\phi}{dt^2} = \frac{1}{t^2} - \frac{1}{(t-1)^2}, \quad \frac{d^2\phi}{dt^2} = 0 \text{ for } t=t_0,$$

$$\frac{d^3\phi}{dt^3} = -\frac{2}{t^3} + \frac{2}{(t-1)^3}, \quad \frac{d^3\phi}{dt^3} = -32 \text{ for } t=t_0.$$

Thus

$$\epsilon^\phi = \epsilon^{\phi_0 + \frac{1}{6} \left(\frac{d^3\phi}{dt^3} \right)_0 (t-t_0)^3 + \dots}$$

$$\phi_0 = 2,$$

and the limit becomes

$$\lim F = \frac{(-1)^{n_1}}{2\pi i} \int_c t^{n_1} (1-t)^{-n_1-1} \epsilon^{2il} e^{-\frac{16}{3} i(t-t_0)^3 + \dots} dt.$$

Because of the large value of l , the decrease with increasing $(t-t_0)$ will be extremely rapid, so that we can regard the remaining factor as constant with the value for $t=t_0=\frac{1}{2}$, i. e. 2. Therefore,

$$\lim F = \frac{(-1)^{n_1}}{\pi i} e^{2il} \int_c e^{-\frac{16}{3} i(t-t_0)^3} dt.$$

Put

$$t-t_0=\delta; \quad \delta=\rho\epsilon^{i\psi}; \quad \delta^3=\rho^3\epsilon^{3i\psi}.$$

δ^3 has positive values when $3\psi=\frac{2\pi n}{3}$; $\psi=2\pi n$. We get, therefore,

three directions, $\psi=0$, $\psi=\frac{2\pi}{3}$, $\psi=\frac{4\pi}{3}$. We conduct the loop around the

origin as indicated in fig. 1.

Only the paths $\overline{BO'}$ $\overline{O'A}$ will give a contribution.

For $\overline{BO'}$
$$dt=d\delta=+d\rho\epsilon^{\frac{i4\pi}{3}}=d\rho\left(-\frac{1}{2}-i\frac{\sqrt{3}}{2}\right)$$

For $\overline{O'A}$
$$dt=d\delta=+d\rho\epsilon^{\frac{i2\pi}{3}}=d\rho\left(-\frac{1}{2}+i\frac{\sqrt{3}}{2}\right).$$

Hence

$$\begin{aligned}\lim F &= \frac{(-1)^{n_1}}{\pi i} \epsilon^{2l} \left[\int_{+\infty}^0 \epsilon^{-\frac{16}{3} l \delta^3} d\delta \left(-\frac{1}{2} - i \frac{\sqrt{3}}{2} \right) \right. \\ &\quad \left. + \int_0^{\infty} \epsilon^{-\frac{16}{3} l \delta^3} d\delta \left(-\frac{1}{2} + i \frac{\sqrt{3}}{2} \right) \right] \\ &= \frac{(-1)^{n_1} \epsilon^{2l}}{\pi} \sqrt{3} \int_0^{\infty} \epsilon^{-\frac{16}{3} l \delta^3} d\delta \\ &= \frac{(-1)^{n_1} \epsilon^{2l}}{\pi} \frac{\sqrt{3}}{3} \left(\frac{16l}{3} \right)^{-\frac{1}{3}} \Gamma\left(\frac{1}{3}\right).\end{aligned}$$

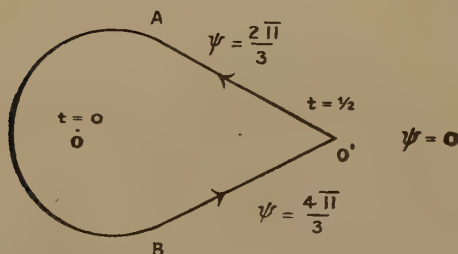
Hence

$$\lim lF = \frac{(-1)^{n_1}}{\pi} \frac{\sqrt{3}}{3} \epsilon^{2l} \left(\frac{16l}{3} \right)^{-\frac{1}{3}} \Gamma\left(\frac{1}{3}\right).$$

The sum of the matrix amplitudes squared is then

$$|J|^2 = \frac{2^7 \pi \mu a^6}{3h} n^2 l^6 \left(\frac{16l}{3} \right)^{-\frac{1}{3}} \overline{\Gamma\left(\frac{1}{3}\right)}^2,$$

Fig. 1.



and the complete intensity

$$I = \frac{2^5 \pi e^6 Z^2}{3^2 \mu h c^3} \cdot \frac{n^2}{l^2} \left(\frac{16l}{3} \right)^{-\frac{1}{3}} \overline{\Gamma\left(\frac{1}{3}\right)}^2.$$

The probability of spontaneous transition is

$$A_{nl} = \frac{2^4 e^2 \hbar}{3^2 \pi c^3 \mu^2} n^2 \left(\frac{16l}{3} \right)^{-\frac{1}{3}} \overline{\Gamma\left(\frac{1}{3}\right)}^2,$$

and the coefficient of absorption per atom

$$\alpha_v = \frac{\hbar^7}{2 \cdot 3^2 \pi^6 e^6 \mu^4 Z^4 c} n^2 l^4 \left(\frac{16l}{3} \right)^{-\frac{1}{3}} \overline{\Gamma\left(\frac{1}{3}\right)}^2.$$

(b) If we consider the more general case when n and l are both large, z is then moderately sized.

$$\begin{aligned}F(-in, -n_2, 1, z) &= \frac{(-1)^{n_1}}{2\pi i} \int_0^1 t^{-n_1-1} (1-t)^{n_2} (1-zt)^{in} dt. \\ &= \frac{(-1)^{n_1}}{2\pi i} \int_0^1 -l(1-t)^{l-1} (1-zt)^{in} dt.\end{aligned}$$

Put

$$\begin{aligned} e^{\phi} &= t^{-l}(1-t)^l(1-zt)^{in}, \\ \phi &= -l \log t + l \log (1-t) + in \log (1-zt), \\ \frac{d\phi}{dt} &= -\frac{l}{t} + \frac{l}{t-1} + \frac{inz}{zt-1}. \end{aligned}$$

For the saddle point $\frac{d\phi}{dt}=0$ we have

$$\begin{aligned} inzt^2 + (l-in)zt - l &= 0, \\ t_{1,2} &= \frac{(in-l)}{2in} \pm \frac{(in-l)}{2in} \sqrt{1 + \frac{4inl}{(in-l)^2} \frac{1}{z}}. \end{aligned}$$

If we use the value $z = \frac{4inl}{(n+il)^2}$ we find

$$1 + \frac{4inl}{(in-l)^2} \frac{1}{z} = 0,$$

so that the saddle point is

$$\begin{aligned} t_1 &= \frac{in-l}{2in} = \frac{n+il}{2n}, \\ \frac{d^2\phi}{dt^2} &= \frac{l}{t^2} - \frac{l}{(t-1)^2} - \frac{inz^2}{(zt-1)^2} \\ &= 0 \text{ for } t=t_1 = \frac{n+il}{2n}. \end{aligned}$$

Again,

$$\begin{aligned} \frac{d^3\phi}{dt^3} &= -\frac{2l}{t^3} + \frac{2l}{(t-1)^3} + \frac{2inz^3}{(zt-1)^3} \\ &= -\frac{2^5ln^4}{(n^2+l^2)^2} \text{ for } t=t_1 = \frac{n+il}{2n}. \end{aligned}$$

Thus

$$\begin{aligned} \lim F &= \frac{(-1)^{n_1}}{2\pi i} \frac{\epsilon^{\phi_1}}{(1-t_1)} \int \epsilon^{\frac{1}{6} \left(\frac{d^3\phi}{dt^3} \right) (t-t_1)^3 + \dots} dt, \\ \phi_1 &= \log \left(\frac{n-il}{n+il} \right)^{l+in} \\ &= (l+in) \left(-2i \arctan \frac{l}{n} \right) \\ &= 2(n-il) \arctan \frac{l}{n}. \\ 1-t_1 &= \frac{n-il}{2n}. \end{aligned}$$

$$\text{Therefore } \lim F = \frac{(-1)^{n_1}}{2\pi i} \frac{2n}{(n-il)} \epsilon^{2(n-il) \arctan \frac{l}{n}} \int \epsilon^{-\frac{16}{3} \frac{ln^4}{(n^2+l^2)^2} (t-t_1)^3} dt.$$

Because of the large value of $\frac{ln^4}{(n^2+l^2)^2}$ the decrease with increasing $(t-t_1)$ will be extremely rapid, so that the remaining factor has been regarded as constant with the value for $t=t_1 = \frac{n+il}{2n}$.

If we put $t-t_1=\delta=\rho e^{i\psi}$; $\delta^3=\rho^3 e^{3i\psi}$, then δ^3 will again remain positive for three directions, $\psi=0, \frac{2\pi}{3}$ and $\frac{4\pi}{3}$. If l and n pass to large values in such a way that $\frac{l}{n} \leq \sqrt{3}$, we may conduct the loop around the origin in one of the ways indicated in figs. 2 and 3.

Both ways give the same result, so that we need for the present only consider the evaluation for the first loop.

Fig. 2.

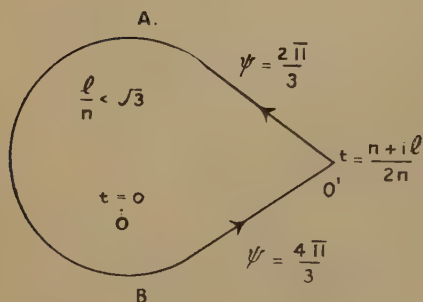
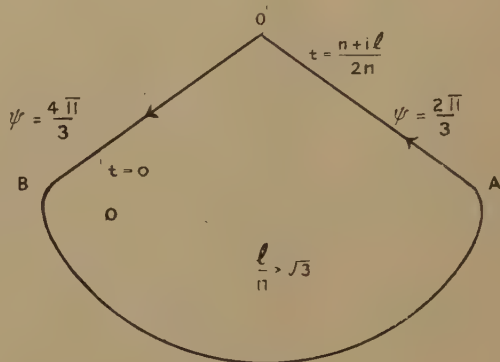


Fig. 3.



Only the paths $\overline{BO'}$ $\overline{O'A}$ will give a contribution, and in a similar way to the preceding evaluation

$$\begin{aligned} \lim F &= \frac{(-1)^{n_s}}{\pi i} \frac{n}{n-il} \epsilon^{2(n-il) \arctan \frac{l}{n}} \left[\int_{+\infty}^0 \epsilon^{-\frac{16}{3} \frac{ln^4}{(n^2+l^2)^2} \delta^3} d\delta \left(-\frac{1}{2} - \frac{i\sqrt{3}}{2} \right) \right. \\ &\quad \left. + \int_0^{\infty} \epsilon^{-\frac{16}{3} \frac{ln^4}{(n^2+l^2)^2} \delta^3} d\delta \left(-\frac{1}{2} + \frac{i\sqrt{3}}{2} \right) \right] \\ &= \frac{(-1)^{n_s}}{\pi} \frac{n}{(n-il)} \epsilon^{2(n-il) \arctan \frac{l}{n}} \sqrt{3} \int_0^{\infty} \epsilon^{-\frac{16}{3} \frac{ln^4}{(n^2+l^2)^2} \delta^3} d\delta \\ &= \frac{(-1)^{n_s}}{\pi} \frac{n}{(n-il)} \epsilon^{2(n-il) \arctan \frac{l}{n}} \frac{\sqrt{3}}{3} \left(\frac{16ln^4}{3(n^2+l^2)^2} \right)^{-\frac{1}{3}} \Gamma\left(\frac{1}{3}\right). \end{aligned}$$

$$\text{Hence } \lim lF = \frac{(-1)^{n_s}}{\pi} \frac{n\dot{l}}{(n-il)} \epsilon^{2(n-il) \arctan \frac{l}{n}} \frac{\sqrt{3}}{3} \left(\frac{16ln^4}{3(n^2+l^2)^2} \right)^{-\frac{1}{3}} \Gamma\left(\frac{1}{3}\right).$$

The sum of the matrix amplitudes squared becomes

$$|J|^2 = \frac{2^7 \pi \mu a^6}{3h} \frac{n^{12} l^6}{(n^2+l^2)^5} \left[\frac{16ln^4}{3(n^2+l^2)^2} \right]^{-2/3} \Gamma\left(\frac{1}{3}\right)^2,$$

and the complete intensity is

$$I = \frac{2^5 \pi e^6 Z^2}{3^2 \mu h c^3} \frac{n^4}{l^2 (n^2+l^2)} \left[\frac{16ln^4}{3(n^2+l^2)^2} \right]^{-2/3} \Gamma\left(\frac{1}{3}\right)^2.$$

The probability of spontaneous transition is

$$A_{nl} = \frac{2^4 e^2 \hbar}{3^2 \pi c^3 \mu^2} \frac{n^6}{(n^2+l^2)^2} \left[\frac{16ln^4}{3(n^2+l^2)^2} \right]^{-\frac{1}{3}} \Gamma\left(\frac{1}{3}\right)^2,$$

and the coefficient of absorption per atom

$$\alpha_{\nu} = \frac{\hbar^7}{2 \cdot 3^2 \pi^6 e^6 \mu^4 Z^4 c} \frac{n^{10} l^4}{(n^2 + l^2)^4} \left[\frac{16 l n^4}{3(n^2 + l^2)^2} \right]^{-\frac{1}{2}} \frac{1}{\Gamma(\frac{1}{2})}.$$

The sum of the matrix amplitudes squared for slow-moving electrons.

It is to be noticed that it is possible to find an expression in a closed form for the sum of the matrix amplitudes squared for $n \rightarrow \infty$, and an arbitrary l .

From the consideration of the hypergeometric series, or from the representation in an integral form, we easily see that

$$\begin{aligned} F(1-in, -n_2; 2, 1-v^2) &\sim F(-n_2, 2, 4l), \\ F(-in, -n_2; 1, 1-v^2) &\sim F(-n_2, 1, 4l), \\ F(-in, -n_2+1, 1, 1-v^2) &\sim F(-n_2+1, 1, 4l). \end{aligned}$$

Also, the bracket expression in the matrix becomes

$$\begin{aligned} &[l + \nu(n_1 + 1)]F(-n_2, 1, 4l) + n_2 \nu F(-n_2 + 1, 1, 4l) \\ &\sim [l + (l - n_2)]F(-n_2, 1, 4l) + n_2 F(-n_2 + 1, 1, 4l) \\ &\sim 2lF(-n_2, 1, 4l) - n_2 [F(-n_2 + 1, 1, 4l) - F(-n_2, 1, 4l)] \\ &\sim 2lF(-n_2, 1, 4l) - l \frac{d}{dl} F(-n_2, 1, 4l) \end{aligned}$$

by using the property that

$$\frac{d}{dx} F(\alpha, j, x) = \frac{\alpha}{x} [F(\alpha + 1, j, x) - F(\alpha, j, x)].$$

Consequently for a given total quantum number l , the sum of the matrix elements squared becomes

$$\begin{aligned} \sum |x(n; n_1 n_2 \pm 1)|^2 &= \sum |y(n; n_1 n_2 \pm 1)|^2 = \\ &|A(n)|^2 2^3 \pi a^5 l^6 \epsilon^{-4l} \sum_{n_2=0}^{(l-2)} (l - n_2 - 1)(n_2 + 1) |F(-n_2, 2, 4l)|^2, \end{aligned}$$

and

$$\sum |z(n; n_1 n_2 0)|^2 = |A(n)|^2 2^5 \pi a^5 l^6 \epsilon^{-4l} \sum_{n_2=0}^{(l-1)} |2F(-n_2, 1, 4l) - \frac{d}{dl} F(-n_2, 1, 4l)|^2.$$

Let us consider the summation

$$\sum (l - m - 1)(m + 1) |F(-m, 2, s)|^2.$$

We have the formulæ

$$F(-m, 2, s) = \frac{\epsilon^s}{s} \frac{1}{[m+1]} \frac{d^m}{ds^m} (\epsilon^{-s} s^{m+1}) = (-1)^m [m] T_1^m(s),$$

where $T_1^m(s)$ is a Sonine Polynomial. If we integrate by parts m times we have the relation

$$(-1)^m \int_0^\infty \epsilon^{-xs} T_1^m(s) \frac{d^m}{ds^m} [\epsilon^{-s} s^{m+1}] ds = \int_0^\infty \epsilon^{-s} s^{m+1} \frac{d^m}{ds^m} [\epsilon^{-xs} T_1^m(s)] ds.$$

If we use the property that

$$\frac{d^p}{ds^p} [T_m^n(s)] = T_{m+p}^{n-p}(s),$$

we find that

$$\begin{aligned} \frac{d^m}{ds^m} [\epsilon^{-xs} T_1^m(s)] &= \epsilon^{-xs} \left[T_{m+1}^0(s) - \frac{m}{1} x T_m^1(s) + \dots \right. \\ &\quad \left. + (-1)^r \frac{m(m-1) \dots (m-r+1)}{r!} x^r T_{m-r+1}^r(s) + \dots \right. \\ &\quad \left. + (-1)^m x^m T_1^m(s) \right] \quad (m+1) \text{ terms.} \end{aligned}$$

Hence the integral may be written in the following form :—

$$\begin{aligned} \int_0^\infty \epsilon^{-s(x+1)} s^{m+1} \left[T_{m+1}^0(s) - \frac{m}{1} x T_m^1(s) + \dots \right. \\ \left. + (-1)^r \frac{m(m-1) \dots (m-r+1)}{r!} x^r T_{m-r+1}^r(s) + \dots \right. \\ \left. + (-1)^m x^m T_1^m(s) \right] ds. \end{aligned}$$

But Sonine ⁽⁶⁾ gives the formula

$$\int_0^\infty \epsilon^{-s(x+1)} s^m T_m^n(s) ds = \frac{(-1)^n x^n}{n(1+x)^{m+n+1}}.$$

Therefore

$$\int_0^\infty \epsilon^{-s(x+1)} s^{m-r+1} T_{m-r+1}^r(s) ds = \frac{(-1)^r x^r}{r(1+x)^{m+2}}.$$

Differentiating, with respect to x , r times we find

$$\int_0^\infty \epsilon^{-s(x+1)} s^{m-r+1} T_{m-r+1}^r(s) ds = \frac{1}{r} \frac{d^r}{dx^r} \left[\frac{x^r}{(1+x)^{m+2}} \right].$$

Consequently the original integral may be written in the form

$$\begin{aligned} &(-1)^m \int_0^\infty \epsilon^{-xs} T_1^m(s) \frac{d^m}{ds^m} [\epsilon^{-s} s^{m+1}] ds \\ &= \frac{1}{(1+x)^{m+2}} - \frac{m}{1} \frac{x}{1} \frac{d}{dx} \left[\frac{x}{(1+x)^{m+2}} \right] + \dots \\ &\quad + (-1)^r m \frac{(m-1) \dots (m-r+1)}{r!} \frac{x^r}{r} \frac{d^r}{dx^r} \left[\frac{x^r}{(1+x)^{m+2}} \right] \\ &\quad + \dots + (-1)^m \frac{x^m}{m!} \frac{d^m}{dx^m} \left[\frac{x^m}{(1+x)^{m+2}} \right]. \end{aligned}$$

But by Lagrange's Theorem

$$\frac{1}{r} \frac{d^r}{dx^r} \left[\frac{x^r}{(1+x)^{m+2}} \right] = \text{coefficient of } y^r \text{ in } \frac{(1-y)^{m+1}}{(1+x-y)^{m+2}}.$$

Therefore

$$(-1)^m \int_0^\infty \epsilon^{-sx} T_1^m(s) \frac{d^m}{ds^m} [\epsilon^{-s} s^{m+1}] ds = \text{coefficient without } y \text{ in}$$

$$\frac{(1-y)^{m+1}}{(1+x-y)^{m+2}} \left(1 - \frac{x}{y}\right)^m.$$

Hence

$$(-1)^m T_1^m(s) \frac{d^m}{ds^m} [\epsilon^{-s} s^{m+1}] = \text{coefficient without } y \text{ in}$$

$$\frac{1}{2\pi i} \int_c \epsilon^{sx} \frac{(1-y)^{m+1} \left(1 - \frac{x}{y}\right)^m}{(1+x-y)^{m+2}} dx,$$

where C is a simple contour enclosing the origin.

But $(-1)^m T_1^m(s) \frac{d^m}{ds^m} [\epsilon^{-s} s^{m+1}] = (m+1) \epsilon^{-s} s [F(-m, 2, s)]^2.$

Hence $(m+1)[F(-m, 2, s)]^2 = \text{coefficient without } y \text{ in}$

$$\frac{1}{2\pi i} \frac{\epsilon^s}{s} \int_c \epsilon^{sx} \frac{(1-y)^{m+1} \left(1 - \frac{x}{y}\right)^m}{(1+x-y)^{m+2}} dx.$$

Therefore

$$S_1 = \sum_{n_2=0}^{l-2} (l-n_2-1)(n_2+1)[F(-n_2, 2, 4l)]^2$$

$$= \text{coefficient without } y \text{ in}$$

$$\frac{1}{2\pi i} \frac{\epsilon^{4l}}{4l} \int_c \epsilon^{4lx} \frac{(1-y)}{(1+x-y)^2} \sum_{n_2=0}^{l-1} (l-n_2-1) \frac{\left(1 - \frac{x}{y}\right)^{n_2} (1-y)^{n_2}}{(1+x-y)^{n_2}}$$

$$= \text{coefficient without } y \text{ in}$$

$$\frac{1}{2\pi i} \frac{\epsilon^{4l}}{4l} \int_c \epsilon^{4lx} \frac{dx}{x^2} \frac{(1-y)^{l+1} (y-x)^l}{y^{l-2} (1+x-y)^l}.$$

We have now to consider the sum

$$S_2 = \sum_{n_2=0}^{l-1} \left| 2F(-n_2, 1, 4l) - \frac{d}{dl} F(-n_2, 1, 4l) \right|^2$$

$$= 4 \sum_{m=0}^{l-1} \left| F(-m, 1, s) - 2 \frac{d}{ds} F(-m, 1, s) \right|^2.$$

But $\frac{d}{ds} F(-m, 1, s) = -mF(-m+1, 2, s),$

so that $S_2 = 4 \sum_{m=0}^{l-1} [F(-m, 1, s) + 2mF(-m+1, 2, s)]^2$

$$= 4 \sum_{m=0}^{l-1} [F(-m, 1, s)]^2 + 4mF(-m, 1, s)F(-m+1, 2, s)$$

$$+ 4m^2[F(-m+1, 2, s)]^2.$$

From the preceding discussion we note immediately that

$$m[F(-m+1, 2, s)]^2 = \text{coefficient without } y \text{ in}$$

$$\frac{1}{2\pi i} \frac{\epsilon^s}{s} \int_c \epsilon^{sx} \frac{(1-y)^m \left(1 - \frac{x}{y}\right)^{m-1}}{(1+x-y)^{m+1}} dx.$$

In a similar manner, by using the relations

$$F(-m, 1, s) = \frac{\epsilon^s}{[m]} \frac{d^m}{ds^m} [\epsilon^{-s} s^m] = (-1)^m [m] T_0^m(s),$$

and considering the integral

$$(-1)^m \int_0^\infty \epsilon^{-sx} T_0^m(s) \frac{d^m}{ds^m} [\epsilon^{-s} s^m] ds,$$

we find that

$$[F(-m, 1, s)]^2 = \text{coefficient without } y \text{ in}$$

$$\frac{1}{2\pi i} \epsilon^s \int_c \epsilon^{+sx} \frac{(1-y)^m \left(1 - \frac{x}{y}\right)^m}{(1+x-y)^{m+1}}.$$

Also, by using the relations

$$F(-m+1, 2, s) = \frac{\epsilon^s}{s} \frac{1}{[m]} \frac{d^{m-1}}{ds^{m-1}} [\epsilon^{-s} s^m] = (-1)^{m-1} [m-1] T_1^{m-1}(s),$$

$$F(-m, 1, s) = \epsilon^s \frac{1}{[m]} \frac{d^m}{ds^m} [\epsilon^{-s} s^m] = (-1)^m [m] T_0^m(s),$$

and considering the integral

$$(-1)^{m-1} \int_0^\infty \epsilon^{-sx} T_0^m(s) \frac{d^{m-1}}{ds^{m-1}} [\epsilon^{-s} s^m] ds,$$

we find that

$$F(-m, 1, s) F(-m+1, 2, s) = \text{coefficient without } y \text{ in}$$

$$-\frac{1}{2\pi i} \frac{\epsilon^s}{s} \int_c \epsilon^{sx} \frac{(1-y)^m \left(1 - \frac{x}{y}\right)^{m-1}}{y(1+x-y)^{m+1}} dx.$$

Therefore,

$$S_2 = 4 \sum_{m=0}^{l-1} [F(-m, 1, s)]^2 + 4 \sum_{m=0}^{l-1} 4m [F(-m, 1, s) F(-m+1, 2, s) + m \{F(-m+1, 2, s)\}^2]$$

= coefficient without y in

$$\frac{2}{\pi i} \epsilon^s \int_c \frac{\epsilon^{sx}}{(1+x-y)} \sum_0^{l-1} \lambda^m - \frac{8}{\pi i} \frac{\epsilon^s}{s} \int_c \frac{\epsilon^{sx} (1-y)}{y \left(1 - \frac{x}{y}\right) (1+x-y)} \sum_0^{l-1} m \lambda^m,$$

where

$$\lambda = \frac{(1-y) \left(1 - \frac{x}{y}\right)}{(1+x-y)}.$$

Hence S_2 = coefficient without y in

$$\begin{aligned} & \frac{2}{\pi i} \epsilon^s \int_c \frac{\epsilon^{sx} dx}{(1+x-y)} \left(\frac{1-\lambda^l}{1-\lambda} \right) - \frac{8}{\pi i} \frac{\epsilon^s}{s} \int_c \epsilon^{sx} \frac{(1-y)^2 dx}{y(1+x-y)^2} \frac{d}{d\lambda} \left(\frac{1-\lambda^l}{1-\lambda} \right) \\ & = \text{coefficient without } y \text{ in} \\ & \frac{2}{\pi i} \epsilon^{4l} \int_c \frac{\epsilon^{4lx} dx}{(1+x-y)} \left(\frac{1-\lambda^l}{1-\lambda} \right) - \frac{2}{\pi i} \frac{\epsilon^{4l}}{l} \int_c \epsilon^{4lx} \frac{(1-y)^2 dx}{y(1+x-y)^2} \frac{d}{d\lambda} \left(\frac{1-\lambda^l}{1-\lambda} \right). \end{aligned}$$

The sum of the matrix amplitudes squared is now determined in a closed form.

References.

- (1) Temple, Proc. Roy. Soc. cxxi. p. 673 (1928).
- (2) Temple, *loc. cit.*
- (3) Oppenheimer, *Zeits. f. Phys.* p. 725 (1929).
- (4) Epstein, Phys. Rev. vol. xxviii. p. 695 (1926).
- (5) Morgans, Phil. Mag. vol. xiii. p. 664 (1932).
- (6) Sonine, Math. Ann. xvi. p. 42 (1880).

LXIII. Infinitesimal Bending of a Surface.

By LUDWIK SILBERSTEIN, Ph.D.*

[Received January 15, 1943.]

WITH any parameters u, v , let the vector equation of the surface be $\mathbf{r} = \mathbf{r}(u, v)$, and $\mathbf{f} = \mathbf{f}(u, v)$ the infinitesimal displacement of any of its points, such, that is, that f^2 -terms are to be neglected. The directed line-element of the deformed surface is $d\mathbf{r}' = d\mathbf{r} + d\mathbf{f}$, and its squared size, up to second-order terms,

$$ds'^2 = (d\mathbf{r})^2 + 2d\mathbf{r} \cdot d\mathbf{f} + ds^2 + 2d\mathbf{r} \cdot d\mathbf{f}.$$

Hence, by the definition of "bending," which claims the conservation of length of any element, $d\mathbf{r} \cdot d\mathbf{f} = 0$ or, if derivatives with respect to u, v are denoted by the subscripts 1, 2,

$$(\mathbf{r}_1 du + \mathbf{r}_2 dv)(\mathbf{f}_1 du + \mathbf{f}_2 dv) = 0.$$

This condition has to be satisfied for all du, dv , so that it splits into the three conditions

$$\mathbf{r}_1 \mathbf{f}_1 = \mathbf{r}_2 \mathbf{f}_2 = \mathbf{r}_1 \mathbf{f}_2 + \mathbf{r}_2 \mathbf{f}_1 = 0. \quad . \quad . \quad . \quad . \quad . \quad (1)$$

These are Weingarten's equations (Crelle's Journal, vol. c.), only written in a compact vector form. Weingarten himself has developed them (*l. c.*) in terms of the Cartesian components of \mathbf{r} and \mathbf{f} in a manner which ultimately leads to a second-order partial differential equation for an auxiliary function, his "displacement function" or the

* Communicated by the Author.

"characteristic" function. If any solution of this equation is found, the corresponding displacement can be determined by quadratures. A discussion of the well-known technical difficulties in actual applications of this method to given surfaces must not detain us here.

The purpose of this paper is to offer an alternative developed form of the equations (1) which has seemed to be more simple and, perhaps, more natural.

Let u, v be an orthogonal system, so that the vectors \mathbf{r}_1 and \mathbf{r}_2 , tangential to the u - and v -lines, are perpendicular to each other, and the line-element of the original surface is

$$ds^2 = \mathbf{r}_1^2 du^2 + \mathbf{r}_2^2 dv^2 \equiv E du^2 + G dv^2.$$

The unit normal of the surface is $\mathbf{n} = \frac{1}{\sqrt{EG}} \nabla \mathbf{r}_1 \mathbf{r}_2$ and \mathbf{r}_1/\sqrt{E} , \mathbf{r}_2/\sqrt{G} are

unit vectors tangential to the surface and normal to each other. It thus seems most natural to split the displacement \mathbf{f} into components along these three vectors, at any point of the surface. If $R, S=s\sqrt{E}$, $T=t\sqrt{G}$ be the normal and the two tangential components, we have

$$\mathbf{f} = R\mathbf{n} + s\mathbf{r}_1 + t\mathbf{r}_2, \quad (2)$$

whence

$$\mathbf{f}_1 = R_1\mathbf{n} + s_1\mathbf{r}_1 + t_1\mathbf{r}_2 + R\mathbf{n}_1 + s\mathbf{r}_{11} + t\mathbf{r}_{12},$$

$$\mathbf{f}_2 = R_2\mathbf{n} + s_2\mathbf{r}_1 + t_2\mathbf{r}_2 + R\mathbf{n}_2 + s\mathbf{r}_{12} + t\mathbf{r}_{22}.$$

Substitute the scalar products of \mathbf{f}_1 and \mathbf{f}_2 into $\mathbf{r}_1, \mathbf{r}_2$ in (1), recalling that $\mathbf{n}\mathbf{r}_1 = \mathbf{n}\mathbf{r}_2 = \mathbf{r}_1\mathbf{r}_2 = 0$, $\mathbf{n}^2 = 1$, $\mathbf{r}_1^2 = E$, $\mathbf{r}_2^2 = G$, whence, *e. g.*, $\mathbf{r}_1\mathbf{r}_{11} = \frac{1}{2}E_1$, $\mathbf{r}_1\mathbf{r}_{22} = -\frac{1}{2}G_1$, etc., $\mathbf{n}_1\mathbf{r}_1 = -\mathbf{n}\mathbf{r}_{11}$, $\mathbf{n}_2\mathbf{r}_2 = -\mathbf{n}\mathbf{r}_{22}$, $\mathbf{r}_1\mathbf{n}_2 + \mathbf{r}_2\mathbf{n}_1 = -2\mathbf{n}\mathbf{r}_{12}$, and using the familiar symbols L, M, N for

$$\mathbf{n}\mathbf{r}_{11}, \mathbf{n}\mathbf{r}_{12}, \mathbf{n}\mathbf{r}_{22},$$

the coefficients of the "second fundamental form" of the surface.

Then the equations for the most general displacement giving an infinitesimal bending of the surface will assume the remarkably simple form

$$\left. \begin{aligned} Es_1 + \frac{1}{2}(E_1s + E_2t) &= LR, \\ Es_2 + Gt_1 &= 2MR, \\ Gt_2 + \frac{1}{2}(G_1s + G_2t) &= NR. \end{aligned} \right\} \quad (3)$$

These are three linear homogeneous equations for the normal component R and the reduced tangential components, $s=S/\sqrt{E}$, $t=T/\sqrt{G}$, of the displacement \mathbf{f} with coefficients E , etc., given functions of u, v . They are of the first order in s, t , but of the zeroth order in R , i. e., contain only R , not its derivatives. In fact, the derivatives R_1, R_2 , which appear in $\mathbf{f}_1, \mathbf{f}_2$ only as coefficients of the normal \mathbf{n} , are knocked out on multiplying with either tangential vector, \mathbf{r}_1 or \mathbf{r}_2 . Thus R can be directly eliminated, leaving us with two linear, homogeneous equations for s, t . The sum of any solutions is again a solution.

Some interesting properties may be read off the equations (3) at a glance. They teach us that (unless the surface be a plane) a purely

normal bending-displacement is impossible. For, if $s=t=0$ throughout, we have also $R=0$ (unless $L=M=N=0$), and thus no displacement at all*. If, on the other hand, we require the displacement to be purely tangential, i. e. $R=0$ or a shifting of the surface in itself, we have for s, t alone three equations which can be compatible with each other only if the coefficients E, G and their derivatives satisfy a certain relation. This relation, when developed, will express a peculiarity of the surface itself, such, e. g., as the constancy of its curvature.

The equations (3), symmetrical in s, t, E, G and their first derivatives, hold for any surface whatever.

For a *surface of revolution* their structure is considerably simplified. The line-element of any such surface may be written $ds^2=du^2+G(u).dv^2$, so that $E=1, E_1=E_2=G_2=0$. Again, as can be readily proved, $M=\mathbf{nr}_{12}=0$. Thus the equations (3) become

$$\frac{\partial s}{\partial u}=LR, \quad \frac{\partial s}{\partial v}G+\frac{\partial t}{\partial u}=0, \quad G\frac{\partial t}{\partial v}+\frac{1}{2}G's=NR, \quad \dots \quad (3r)$$

where, moreover, $s=S$, while $t=T/\sqrt{G(u)}$. Eliminating R and t , we have, for the meridian component of the displacement,

$$\frac{\partial}{\partial u}\left(\frac{N}{GL}\frac{\partial s}{\partial v}\right)+\frac{1}{G}\frac{\partial^2 s}{\partial v^2}=\frac{\partial}{\partial u}(\gamma's), \quad \dots \quad (3'r)$$

where $\gamma=\frac{1}{2}\log G$. These equations hold for any bending deformation of the surface. In the sub-case of an *axially symmetrical* distribution of displacement, i. e., for s, t, R functions of u alone, their solution is at once reduced to a single quadrature. In fact, the second of (3r) gives $t=\text{const.}=\alpha$, say, and the remaining two equations become

$$G's=2NR=\frac{2N}{L}S'.$$

Ultimately, then, we have, as the most general axially symmetrical bending of any surface of revolution,

$$T=\alpha\sqrt{G(u)}, \quad S=\epsilon e^{\frac{1}{2}\int \frac{1}{N}dg}, \quad R=\frac{S'}{L}=\frac{SG'}{2N}, \quad \dots \quad (3rs)$$

where ϵ is any small constant, independent of the constant α . The transversal component T represents merely a rigid rotation of the whole surface (through an angle α) about its axis, and is thus irrelevant. (We may as well take $\alpha=0$.) A certain part of the remaining displacement (S, R) may be just a rigid translation of the surface along its axis†. If this be discarded, the residue only will be a true (bending) deformation of the surface. The more general deformations $s(u, v)$, etc., ruled by (3r), will offer of course a much greater variety of possibilities.

* If $L=M=N=0$ (plane), the distribution of the infinitesimal $R=R(u, v)$ remains perfectly arbitrary. In fact, a purely normal displacement produces only second-order changes in the line-elements, whatever its distribution.

† We shall see presently that, for a sphere, all of it is such a translation.

As a further illustration of the proposed method of treatment of infinitesimal bendings, let us consider the case of a *sphere*, whose bending possibilities and associated properties have occupied the attention of geometers for over a century. We shall see that, in this case, the complete solution of the equations (3r) can be found with comparative ease.

With the radius of the sphere as unit length, u, v as co-latitude and longitude (equivalent to $\mathbf{r} = \mathbf{i} \cos u + \sin u [\mathbf{j} \cos v + \mathbf{k} \sin v]$), its line-element is $du^2 + \sin^2 u \cdot dv^2$. Thus, $G = \sin^2 u$, $G' = \sin 2u$, $t = T/\sin u$, and $s = S$. Again, $L = \mathbf{n}\mathbf{r}_{11} = -1$, $M = \mathbf{n}\mathbf{r}_{12} = 0$, as for every surface of revolution, and $N = \mathbf{n}\mathbf{r}_{22} = -\sin^2 u$. Eliminate the radial component R from the first of (3r),

$$R + \frac{\partial S}{\partial u} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Then the remaining two equations, for S along a meridian and T along a latitude circle, will become

$$\begin{aligned} \sin u \frac{\partial S}{\partial u} - \frac{\partial T}{\partial v} S &= \cos u, \\ \sin u \frac{\partial T}{\partial v} + \frac{\partial S}{\partial v} &= T \cos u. \end{aligned}$$

These equations have the remarkable property that they are transformed into themselves on replacing S by T and T by $-S$, just as Maxwell's electromagnetic equations on replacing \mathbf{E} by \mathbf{M} and \mathbf{M} by $-\mathbf{E}$. The latter property induced me many years ago to consider the "bivector" $\mathbf{E} + i\mathbf{M}$, which has condensed Maxwell's two vector equations into a single one*. The analogous property has now prompted me to consider the complex magnitude $Q = S + iT$. In fact, the last two equations are thus condensed into

$$\sin u \frac{\partial Q}{\partial u} + i \frac{\partial Q}{\partial v} = \cos u \cdot Q.$$

Next, introduce the new variable u' through $\frac{du}{\sin u} = du'$ or $u' = \log \tan \frac{u}{2}$

and, at the same time, $\cos u = -\tanh u'$, $\sin u = \frac{1}{\cosh u'}$. Then the equation will become

$$\frac{\partial Q}{\partial u'} + i \frac{\partial Q}{\partial v} + \tanh u' \cdot Q = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

A particular integral is $Q = \operatorname{sech} u'$. This suggests the form $Q = \operatorname{sech} u' \cdot F(u', v)$, where F is, so far, an unrestricted function of both variables. Now, a substitution in (5) gives

$$\frac{\partial F}{\partial u'} + i \frac{\partial F}{\partial v} = 0,$$

* *Annalen d. Physik*, vol. xxiv. (1907).

whence $F=F(u'+iv)$, any function of its argument. Hence, the complete solution of (5),

$$Q=F.\operatorname{sech} u'=\sin u.F(u'+iv).$$

Ultimately, by (4), and writing $F=X+iY$,

$$\left. \begin{aligned} S &= X.\sin u, & T &= Y.\sin u, \\ R &= -X \cos u - \sin u \frac{\partial X}{\partial u}, \end{aligned} \right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $X+iY$ is any function of $u'+iv=\log \tan \frac{u}{2}+iv$. This, then, is the most general "infinitesimal bending" of a sphere comprising, through F , a multiple infinity of various displacements. The simplest of these corresponds to $F=\text{const.}=\epsilon+i\alpha$, say, where ϵ and α are independent constants. This gives an axially symmetrical distribution,

$$R=-\epsilon \cos u, \quad S=\epsilon \sin u, \quad T=\alpha \sin u^*, \quad . \quad . \quad . \quad (6a)$$

which, moreover, is readily seen to be the only possible axially symmetrical displacement. It manifestly represents a rigid rotation α of the whole sphere about the axis, paired with its translation $-\epsilon$ along the axis, or a rigid screw motion of the sphere. In fine, $F=\text{const.}$ gives just the most general rigid motion of the whole sphere. No bending. And since axial symmetry (independence of v) implies necessarily $F=\text{const.}$, we see, at the same time, that the sphere *cannot be bent* into any surface of revolution, without or even with singularities (that is to say, neither the sphere as a whole nor any complete zonal portion of the sphere).

There is, of course, a whole family of surfaces of revolution of constant curvature $K=1$, namely, those (and only those) having for their meridian curve

$$x=\int \sqrt{1-k^2 \sin^2 u} . du, \quad y=k \cos u,$$

where u is our previous $\frac{\pi}{2}-u$ and k^2 is any positive constant †, but our sphere ($k^2=1$) cannot be bent into any one of them without, that is, being sliced up. For their equators, *e. g.*, are either shorter or longer than 2π . In fact, start from the sphere, $k=1$, and since k is to be infinitesimal, suppose $k=1+\epsilon$, where ϵ^2 is to be negligible. Then $k^2=1+2\epsilon$. The first part of the line-element, $du^2=dx^2+dy^2$, remains unchanged, but y^2dv^2 produces a first-order change, $2\epsilon \cos^2 u . dv^2$, against the definition of infinitesimal bending, which verifies our last proposition.

* This is a sub-case of (3rs). In fact, for the sphere, $L/N=1/G$, so that $e^{\frac{1}{2}\int \frac{L}{N} dG} = \sqrt{G} = \sin u$, and so on.

† Those with $k^2 < 1$ are spindle-shaped, with sharp points at the ends, and those with $k^2 > 1$ barrel-shaped, with sharp ridges or folds.

This, then, rules out the possibility of axially symmetrical bendings of a sphere. If any others are possible, they must be of the form (6), with $X+iY=F(u'+iv)$ not a constant. There is an embarrassing multitude of forms of F , but it seems very doubtful that any of them should be free from unremovable singularities, due to $u'=\log \tan \frac{u}{2}$. It will be kept in mind that the range of u is zero to π and that of v (say) 0 to $\pm\pi$. Moreover, π and $-\pi$, generally, v and $v+2\pi$, define the same places on the sphere. Thus, *e. g.*, $F=\epsilon \sin(u'+iv)$, which leads to

$$T=\epsilon \frac{\cos u'}{\cosh u'} \cdot \sinh v,$$

would, through the last factor, give a sharp discontinuity all along a half meridian, and so on.

Another simple example is $F=\epsilon e^{u'+iv}$. This gives

$$\left. \begin{aligned} R &= -\epsilon \left\{ \tan \frac{u}{2} + \frac{\sin u}{1+\cos u} \right\} \cos u \cdot \cos v, \\ S &= \epsilon \sin u \cdot \tan \frac{u}{2} \cdot \cos v, \quad T = \epsilon \sin u \tan \frac{u}{2} \cdot \sin v, \end{aligned} \right\} \quad (6.1)$$

$S^2+T^2=\epsilon^2 \sin^2 u \cdot \tan^2 \frac{u}{2} = \epsilon^2 \sin^2 u \frac{1-\cos u}{1+\cos u}$. This displacement is regular (one-valued, continuous, and finite) all over the sphere, except the south pole, $u=\pi$, where $\tan \frac{u}{2}=\infty$. Both S and T contain only the product $\sin u \cdot \tan \frac{u}{2}$ which, at that pole, tends to 1. The radial component, however, contains the bracketed expression, which, with $u=\pi-\theta$ ($\theta \rightarrow 0$) tends to ∞ as $3/\theta$. In fine, S and T are satisfactory all over the sphere, but the radial component R of the displacement becomes infinite at a single point, the "south pole." If a minute cap is cut out around that point (which may, of course, be any point), the sphere can be bent according to (6.1). This would harmonize with the "rigidity of the sphere" as a whole, guessed at by Minding (1838) and rigorously proved by Liebmann (1899) who, twenty years later, added the consolatory clause*: if ever so small a hole is cut into the sphere, the remaining surface can be bent; but, as we saw a moment ago, certainly not in an axially symmetrical way.

The discussion of other possible bending deformations of the sphere, all contained in (6), may be left to the reader.

Rochester, N.Y.

October 4, 1942.

* H. Liebmann, Munich Berichte, 1919. I quote from W. Blaschke's 'Differentialgeometrie,' I., Berlin, 1930, where the bibliography of the subject may be found.

LXIV. *New Contributions to Interferometry. Part I.—New Non-Localized Interference Fringes.*

By S. TOLANSKY, Ph.D., D.I.C., Physics Dept., Manchester University*.

[Received June 18, 1943.]

[Plates I.–III.]

SUMMARY.

New non-localized interference fringes, produced with a small source and a Fabry-Perot interferometer, are described. These extend into space in the form of cones, producing sharp rings when allowed to fall upon a screen. The properties of these rings are examined. A striking characteristic is that without lenses sharp rings of great diameter can be produced (diameter of first ring equal to half a metre). Complex patterns are produced with various slit sources. Possible applications of the new fringes are considered.

THE FORMATION OF THE NON-LOCALIZED FRINGES.

A FABRY-PEROT interferometer is used, where possible, with an extended source, especially when the instrument is placed in the parallel beam. The visible ring system appears only on the area covered by the source image when the latter arrangement is employed, hence the image must extend at least over the fringe system. The Fabry-Perot fringes are at infinity and are projected onto a screen or photographic plate with a lens, the size of the pattern observed being proportional to the focal length of the image-forming lens.

Despite the long history of the use of the Fabry-Perot interferometer, it has not apparently been formerly noticed that a small source (an approximate point source) can lead to the formation in space of *non-localized* fringes which extend in cones from the interferometer outwards, indefinitely. These cones diverge from the small source, and from them circular fringes can readily be formed on a screen. The latter exhibit novel properties which will be considered later. A simple mathematical treatment will be given later to show how these conical fringes arise, but even without any detailed analysis their origin is clear from the following considerations. A property of the Fabry-Perot interferometer (frequently overlooked) is that in effect it behaves as an "angular filter." The positions of the interference maxima are governed by the well-known expression $N\lambda = 2\mu t \cos \phi$ (the symbols having the usual

* Communicated by the Author.

meanings). Because of the multiple beams the fringe maxima are very sharp and, in effect, appreciable light is only transmitted by the instrument in regions close to those values of ϕ for which N is integral. To a crude approximation it can be considered that only the light emerging along the regions close to these integral values of ϕ is strong, light for other angles of emergence being effectively suppressed. The light therefore emerges along cones of a definite thickness depending upon the reflecting coefficient of the interferometer mirrors, providing certain conditions relating to the size of the source and the value of t are fulfilled.

If, now, a screen be placed between the interferometer and infinity, circular rings appear upon it. If a certain minimum distance between source and screen is exceeded, the rings formed very closely resemble Fabry-Perot rings. *No lenses are needed for the production of these rings, since the conical fringes are non-localized*, spreading out indefinitely. The further the screen is removed from the source, the larger are the rings. These rings, formed by the projection of the non-localized conical fringes on to *any* plane, are distinctly different from the normal Fabry-Perot rings, which are at infinity and can only be projected onto a screen with the aid of a lens (or seen by accommodating on infinity). The *appearance* of the two types of rings is, however, very similar, for reasons which will appear in the discussion.

Examples of the rings formed by the green mercury line are shown in Pls. I. and II. These were obtained by simply placing photographic plates at the appropriate distances, the source consisting of a hot vacuum arc before which was an aperture of diameter 1.5 millimetres, the interferometer gap being 1.5 millimetres. No other optical components are required.

THE SIZE OF THE RINGS.

The remaining discussion is, for convenience, restricted to the consideration of observations made on the green mercury line, using a silvered Fabry-Perot interferometer, with plate separation 1.5 millimetres and with a reflecting coefficient of the order of 0.80.

An outstanding feature of the rings formed from the conical non-localized fringes is the ease with which very sharp rings of exceptionally great diameters can be formed, exhibiting at the same time high resolution. It is easy, with an ordinary low current arc, to produce a visible ring system on a screen in which the first ring has the relatively enormous diameter of *half a metre*, higher order outer ring diameters exceeding several metres. The *angular* diameter of the p th ring is $\phi_p = \sqrt{4\lambda p/t}$ if, for convenience, the case is taken at which the order of interference at the centre is integral. The *linear* diameters of the rings on a screen are those which would be formed by the same interferometer used in the normal manner of Fabry and Perot, employing a projecting lens with focal length equal to the source-screen distance in the conical fringe

observation. With the present arrangement the first ring has a diameter of half a metre on a screen situated some 13 metres from the source.

When it is remembered that no lenses at all are employed in the production of these rings, the appearance of such unusually large and yet extremely sharply defined narrow rings of high resolution is indeed remarkable. A direct contact print (*unit magnification*) of the rings formed upon a photographic plate placed some 2.75 metres from the source is shown in Pl. I. To take into account the reduction in printing reproduction, a centimetre scale drawn alongside enables the actual dimensions to be appreciated. The size of the complete ring system that could be photographed was, in fact, restricted entirely by the size of photographic plate available (10 in. \times 8 in.). The source was an ordinary hot vacuum mercury vapour lamp used with a green filter, the dimensions being restricted by placing before the arc a metal sheet pierced with a hole of 1.5 millimetre cross-section.

In Pl. II. A, Pl. II. B, and Pl. II. C, are shown the appearances of the rings with source-screen distances respectively 1.75, 2.75 and 4.5 metres. It can be seen that in Pl. II. C the first ring has a diameter of some 16 centimetres. Pl. II. B has been taken with the interferometer spacing very slightly altered, to bring the first visible order nearer the centre, to a position of greater dispersion (see discussion below).

THE SHARPNESS OF THE RINGS.

The observed rings are extremely sharp, and with the source-screen distances adopted here they cannot readily be distinguished (except in extremely high orders) from the corresponding Fabry-Perot rings at infinity. The weak single fringe marked with an arrow in Pl. II. A shows how intrinsically sharp the fringes can be. In fact, the actual width recorded for the main fringes is largely attributable to the natural inherent hyperfine structure of the green mercury line. With an interferometer plate separation of only 1.5 millimetres, and with a hot arc source, very incomplete resolution can only be achieved. Pl. II. B, in which the first ring is placed in a position of high dispersion, shows quite clearly the hyperfine structure of the line. This is just that which can be predicted from the employment of this particular source and interferometer when used in the normal method of Fabry and Perot. The instrumental performance is thus much better than it appears to be, the source line width being in this case the factor limiting fringe sharpness.

Pl. II. D shows rings extending up to the hundredth order.

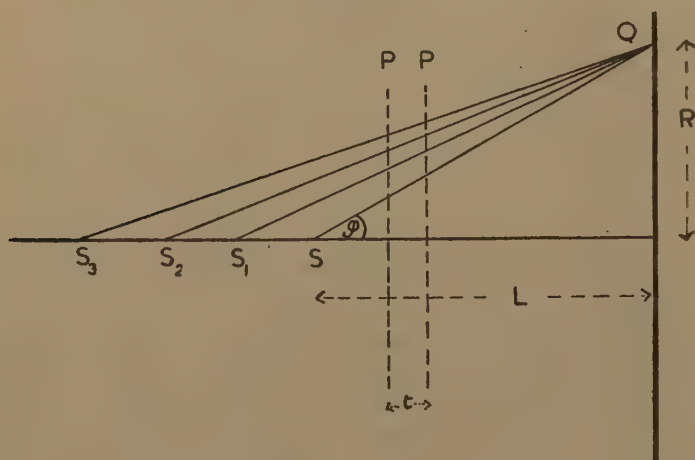
THEORY OF THE FORMATION OF THE FRINGES.

The origin and properties of the fringes are shown by the following simple treatment:—Let it be assumed that a monochromatic point source S is situated at some distance from the interferometer PP, the plate separation of which is t (see fig. 1). As a result of the multiple

reflexions, the effect upon a screen is equivalent to that produced by an infinite number of equidistant point sources S, S_1, S_2, S_3, \dots each $2t$ apart. If the distance L from the source S to the screen is large compared with $2t$, then to a close approximation the successive beams reaching any point Q fall off geometrically in intensity. (If the actual reduction in intensity due to increasing distance be taken into account, the net final effect can be considered as equivalent to a slight reduction in the reflecting coefficient of the mirrors. This is a refinement which need not be considered here.)

It will be observed that the position of the interferometer relative to the source S is immaterial, provided the aperture of the plates suffices to permit the passage of sufficient beams.

Fig 1.



Light from the succession of virtual point sources leads to the formation of circular rings upon the screen distant L from S . (It may be pointed out that a corresponding infinite set of virtual images is formed on the other side of the interferometer. These will lead to the formation of a complementary system of conical non-localized fringes expanding outwards to the left of the diagram. These back reflected fringes are not considered here.) If the lengths of the paths from the successive points S, S_1, S_2, S_3, \dots were to increase in strict arithmetical progression, then it is clear that the rings formed on the screen would have the intensity distribution characteristic of the normal Fabry-Perot fringes. It is shown below that under suitable conditions the deviations of the retardations from this exact condition are not serious.

To a sufficiently close approximation, the path difference at the point Q on the screen of two rays coming from S and S_n respectively equals

$$2nt \cos \phi + \frac{2n^2 t^2}{L} \cdot \cos \phi \cdot \sin \phi.$$

If Q is taken to be a point upon the *first* ring, this expression may be written sufficiently closely

$$2nt \left(1 + \frac{n\lambda}{L} \right) \cos \phi,$$

or alternatively

$$2nt \cos \phi + n^2 \cdot \frac{2t\lambda}{L}.$$

The amount by which the n th ray lags behind the arithmetical progression value needed to lead to perfect Fabry-Perot fringes is thus $(n^2-1) \frac{2t\lambda}{L}$. Selecting the values $t=0.15$ centimetre and $L=3$ metres, leads to a *phase* retardation $\left\{ (n^2-1) \frac{2t}{L} \right\}$ of only 3×10^{-3} of a wave for the first pair. Thus, at the point in question, there is almost complete reinforcement for the earlier and more important members of the series, on a screen situated at the above distance. A retardation of a complete half wave-length is only reached first by the beam from the 23rd image, at which the intensity has fallen to the extremely low value of 1/30,000th that of the first beam.

The extent of the phase retardation depends upon the value of $\frac{2t}{L}$. In the above it equals 1/1000, and for this small value it is clear that the intensity distribution in the neighbourhood of rings not too far from the centre of the system closely approximates to that of the corresponding true Fabry-Perot fringes for the same interferometer.

An alternative procedure is to consider the positions of complete reinforcement of the successive beams, and it will be seen from what follows that this leads to an identical conclusion. The points upon the screen at which the successive beams exactly reinforce (integral wave retardation) do not coincide. It is easy to show that if R is the radius of the *first* ring resulting from the sources S and S₁, then the first ring formed by the beams from S and S_n has a slightly different radius, the difference being $dR=(n-1)\sqrt{\lambda t}$. This quantity is equal to $(n-1)/800$ of the interval between the first and the second ring.

To take an extreme case consider the 40th beam, the intensity of which has fallen to the very low value of 1.7×10^{-8} that of the first beam. The point at which this unites with the first beam is still within 1/20th of an order interval from the point at which the first and second beams combine. Now it can be shown that even with perfectly monochromatic light, the intensity distribution in the Fabry-Perot fringe system with a reflecting coefficient 0.8 is such that the instrumental fringe width exceeds 1/20th of the interval between orders. In practice it is in the neighbourhood of 1/18th of an order. When the minute contribution of the 40th beam is remembered, it is clear that *all the effective beams* fall into a region quite narrow compared with the theoretical fringe width for the equivalent Fabry-Perot interferometer. This means that

the intensity distribution (providing $\frac{2t}{L}$ is small) is practically identical with that of Fabry-Perot fringes.

The simple calculations just made refer to the reinforcement of the successive beams with the first beam only. It is necessary for completion of the argument to consider the points of reinforcement of cases such as S_{n+1} , S_{n+2} , S_{n+3} . . . with S_n , etc. In general, rays from S_n and S_{n+1} reinforce at a point distant $\Delta R = 2n\sqrt{\lambda t}$ from that at which S and S_1 unite. This is of the same order as the displacement dR formerly considered. From the known reflecting coefficient it is possible to plot the total distribution graphically. This shows that in theory the resulting fringes are unsymmetrical, degrading outwards. Actually, however, the distribution of intensity so closely approximates the Fabry-Perot distribution, that differences can only be detected in rings extremely far from the centre. These differences will be surveyed later.

THE DIAMETER FORMULA FOR THE RINGS.

The diameters of the rings are not quite the same as those of the equivalent Fabry-Perot rings, although so close to the latter that differences cannot be easily detected. From expressions already given, it is seen that in effect t in the Fabry-Perot formula has changed to $t\left(1 + \frac{n\lambda}{L}\right)$ for the non-localized beams leading to the formation of the first ring. The quantity $\frac{\lambda}{L}$ is some 1.8×10^{-7} . A graphical construction of the distribution of the points of reinforcement of the more important beams shows that the "optical centre of gravity" of the final resulting fringe is not far from the position of reinforcement of the first two beams S and S_1 . The construction, indeed, shows that for a reflecting coefficient of 0.8 the optical centre of gravity lies in the approximate neighbourhood of $n=5$. It follows then that for the first ring t is effectively increased by only one part in a million.

The quantity $\frac{2nt}{L}$ gives then the increase in order of interference for the first ring compared with that of the Fabry-Perot case. With the values adopted here this has the numerical value 1/200 at the optical centre of gravity. The diameter of the first ring is therefore decreased by 1/200th of an order relative to the Fabry-Perot.

The increase in t is not, however, constant but grows with increasing ring order number ("higher order ring" is a somewhat loose term since the order of interference is a maximum at the centre and diminishes regularly on moving outward. It is to be understood that the term "higher order ring" refers to rings progressively further from the centre, *i. e.* to increase in the value of p .) For the p th ring t must be replaced by $t\left(1 + \frac{pn\lambda}{L}\right)$. Hence the p th ring does not occupy the

position of the corresponding ring in the Fabry-Perot system but is displaced by $p/200$ th of an order inwards. The rings therefore gradually, almost inappreciably, shrink inwards and if superposed on the Fabry-Perot system one ring would be overtaken by the 200th order.

This effect is of little consequence if the rings are to be used for hyperfine structure observations. Firstly, it is usually only necessary to know t to one in a thousand for this purpose. Secondly, in evaluating hyperfine structures the differences in *adjacent* ring diameters are employed (or their squares). The error of only $1/200$ th of an order in moving from one fringe to the next is in many cases not a serious matter. In any event the necessary correction can be applied in any particular case and the error eliminated if so desired.

INCREASE OF FRINGE WIDTH WITH ORDER.

Not only do the diameters of higher order rings change, but in addition a much more obvious effect on the fringe width takes place as p increases. The additional phase retardation is proportional to p being equal to $(n^2-1)p \frac{2t}{L}$ for the n th beam reaching the p th ring. For large values of $p \frac{2t}{L}$ the Fabry-Perot condition is not even approximately fulfilled, and as a result the fringes broaden. Since with the present values $\frac{2t}{L}$ is equal to $1/1000$, the effect due to this is not easily detected even at the 100th ring. In order, therefore, to examine this effect, photographs were taken with L reduced to 70 centimetres, making $\frac{2t}{L} = \frac{1}{233}$.

The predicted onset of broadening was observed. At $p=55$ the lines had broadened out until they resembled Newton's rings in this region, rather than Fabry-Perot rings. After the 55th ring a regular sharpening took place. Such an effect can easily be predicted from simple arithmetical consideration of the lag of beams. The phase retardation for successive beams for the 55th ring is about $\frac{n^2-1}{4}$. Alternate beams are thus a quarter of a wave behind each other. Alternate beams mutually partially interfere instead of reinforcing. As p increases there will be a succession of widely spaced values at which further broadenings will reappear. It was quite impossible to pursue the effect further since this broadening then cannot be separated from effects due to the finite size of the source, which become important when the screen approaches the source, and when p increases. Even with $L=70$ centimetres it was necessary to employ a fine point source of diameter less than half a millimetre.

The simple considerations just developed suffice to give a very close approximation to the true picture. If the amplitudes of all the beams are summed in the usual manner, the resulting intensity distribution can be put in the form

$$I = \frac{1}{r^2} \sum r^{2n} \cdot e^{nA} \cdot e^{n^2B},$$

in which

$$A = 2\pi \cdot \frac{2t}{\lambda} \cos \phi,$$

$$B = 2\pi \cdot \frac{2t}{L} \cos \phi = A \cdot \frac{\lambda}{L},$$

and r is the reflecting coefficient.

This series cannot be easily summed. Dr. Janossy, of Manchester University, has been good enough to examine it for me and gives as an approximate summation the form

$$I \simeq \frac{1}{1 + r^4 - 2r^2 \cos \left\{ A \left(1 + \alpha \frac{\lambda}{L} \right) \right\}}.$$

In this the coefficient α equals

$$r^2 \left\{ \frac{1 - 3r^4 + 2r^2 \cos A}{1 + r^4 - 2r^2 \cos A} \right\},$$

For $r=0.8$ this has a maximum value of about 5. Since αB is small compared with A , the intensity distribution closely approximates to the

true Fabry-Perot distribution, as long as $\frac{2t}{L}$ is small. This more exact

result thus confirms the conclusions which were previously drawn without the aid of an elaborate analysis. Furthermore, it will be noted that the value 5 for α is just that which was formerly obtained by the graphical analysis, so that there is exact numerical agreement between the simple and the more analytical procedures. The simple procedure is perhaps advantageous in that the important physical characteristics are predicted without being obscurely bound up in a formula. This formula agrees with the former treatment, giving those effective changes in equivalent value of t , and in the fringe widths, which were derived from simple arithmetical calculations. If the formula is to be used at all, due regard must be paid to the validity of the approximations made.

THE SHARPNESS OF THE RINGS AND THE SIZE OF THE SOURCE.

As shown by the photograph reproduced in Pl. III. A, if, instead of a single aperture, two well separated apertures are used as the source, there appear two similar independent systems of rings, each originating from a different aperture. The centres of the two systems are separated by an amount equal exactly to the distance between the apertures. (With the flickering arc source employed, one aperture was slightly more illuminated than the other, making one ring system a little stronger.) It is clear from this observation that if the light source is a single disk, each point on the disc produces its own independent ring system. It follows then that the actual width of every ring is increased because of the finite source dimension. To the natural ring width, which depends upon the reflecting coefficient and the order number of the ring, must be added a value equal to the diameter of the source. This is a constant value, to be added to every ring in the whole system.

The actual fringe width due to the instrumental intensity distribution is linearly proportional to the source-screen distance, hence the larger the value of L the less important is the influence of the finite source dimension. With $L=3$ metres the natural fringe width of the first fringe is of the order of 1.5 millimetres. If no appreciable broadening is to result (and it is to be noted that the *emission* line width of the source usually considerably exceeds the theoretical *instrumental* width), then the source diameter should be of the order of 1 millimetre or less. With fringes due to lines broadened either by hyperfine structure, Stark, Doppler, or pressure effects, a larger aperture can safely be tolerated.

Since higher order rings approach each other and become narrower, the number of these rings which can be distinguished diminishes as the source diameter increases and as L decreases. This perhaps accounts for the failure in detecting the non-localized fringes earlier. It has been the general practice to employ an extended source with the Fabry-Perot interferometer. A typical source, say a mercury arc extending over some $2\frac{1}{2}$ centimetres, is so large that the first diffuse ring can barely be discerned on a screen distant 5 metres from the source. On the other hand, with a small opening of the order of some $\frac{3}{4}$ millimetre cross-section placed before the arc, a ring can be clearly seen as near as 12 centimetres from the source.

The detrimental effect of the source is proportional to \sqrt{t} , hence the smaller the plate separation the larger the source that can be tolerated.

If L is large, a relatively big disk source can be employed without causing a serious increase in the relative sharpness of the rings. For example, extremely bright rings which are still quite sharp (although distinctly broadened) can be seen at a distance of several metres from the source if the latter has a diameter of as much as 5 millimetres or even more. It is important to note that the background intensity between fringes is not seriously increased by the employment of a large disk source.

SOME PROPERTIES OF THE FRINGES.

(1) The rings vanish if a piece of ground glass is interposed between the interferometer and the source, since each point on the now extended diffuse source produces a separate ring system and these superpose.

(2) The rings vanish if the incident beam from the point source is rendered parallel with a lens. All the incident light now falls normally on the interferometer, leading to the formation of the centre of the system only.

(3) A lens placed between the interferometer and screen, but close to the former, collects a large number of separate cones and thus projects a bright complete system of true Fabry-Perot rings at its focal plane. If the lens be moved *away* from the interferometer so as to collect but a few cones, only a few fringes are projected. If it is displaced laterally off the optic axis, only the few fragments of cones which fill the lens

lead to the projection of the corresponding few fragments of Fabry-Perot rings.

(4) The rings formed from the non-localized conical fringes are not affected by the position of the interferometer, providing this is not so far from the source as to fail to collect the cone corresponding to the highest order under observation. For example, with plates of 5 centimetres diameter the interferometer can be placed anywhere over a range of 45 centimetres if observation is to be restricted to the first nine rings. Over this range the instrument behaves simply as a non-variable angular filter.

NON-CIRCULAR FRINGES.

It is easy to produce an extensive variety of non-circular fringe patterns upon a screen by employing suitably restricted sources. Three typical examples will be briefly described. They are illustrated in Pl. III.

(1) Pl. III. B.—A narrow horizontal slit produces fringes which, over part of the field, approximate to straight lines. Each point on the slit forms identical rings. The envelopes of these lead to the formation of straight lines, as long as the slit, the separations obeying the usual parabolic law of the rings. In other regions the rings tend to overlap to form a continuum.

(2) Pl. III. C.—A source consisting of two slits crossing at right-angles produces two similar perpendicular slit patterns. These combine to produce fringes, the inner ones being square-shaped, as illustrated.

(3) Pl. III. D.—Three slits meeting at a point to form a letter Y produce a complex interesting pattern, consisting of triangles with curved sides.

The appearance in each case depends upon the value of the source-screen distance.

All the observed patterns can be arrived at by simple geometrical construction of superposition of rings. They are not due to true interference.

POSSIBLE APPLICATIONS OF THE CIRCULAR FRINGES.

(1) The smaller the interferometer gap the larger the number of rings discernible. Good rings can be obtained with interferometer separations up to 6 millimetres, the resolving power then being of the order of half a million. There is no doubt that for small values of $\frac{2t}{L}$ the rings can be used for the detection and evaluation of hyperfine structures. The equations given here are sufficiently accurate for this purpose. The procedure as it stands can only be used on single lines or sources which can be filtered.

(2) The rings are *perfect circles* since no lenses are employed. Lens distortion is of some considerable importance in the Fabry-Perot interferometer used in the normal manner.

(3) The absence of necessary lenses leads to a flexibility in dispersion and magnification which is unique in the field of high resolution instruments. This is of considerable interest.

(4) Such large dispersions are readily available that all the complex errors associated with photographic plate grain and plate resolution can be easily entirely eliminated.

(5) Excellent definition should be obtainable in the ultra-violet region, expensive quartz lenses being dispensed with. Furthermore, the deviations from ring perfection are proportional to λ , hence the defects in the ultra-violet become progressively less, the shorter the wave-length.

(6) The dispersion is so great that it is relatively easy to select a suitable fringe size for the purpose of producing highly monochromatic radiation, with very simple means. So large are the ring diameters that the masking off of a given ring, or even portion of a ring thickness, is a simple matter. This production of a high degree of monochromatism might have applications.

(7) Even single isotopic components of a hyperfine structure pattern can be isolated easily by masking.

(8) The rings afford a very striking laboratory or lecture demonstration. They are much sharper and on an exceedingly greater scale than other non-localized fringes and little skill is needed to produce them. At the same time the fringes exhibit high resolving power as well as high dispersion.

I wish to thank my colleague, Dr. Janossy, for helpful discussion concerning the summation of the series.

LXV. Graphical Differentiation and Integration.

By C. W. HANSEL, Bedford School *.

[Received January 15, 1943.]

THE nomographic numerical evaluation of integrals and differential coefficients has been discussed by the author in *Phil. Mag.* vol. xxxiv. p. 1, Jan. 1943, in an article "An Extension of Nomography." The author has also treated derived curves in an article "Standard Curves," *Phil. Mag.* vol. xxxiv. p. 361, June, 1943. The present paper describes methods supplementary or alternative to those already described.

* Communicated by the Author.

1. Notation.

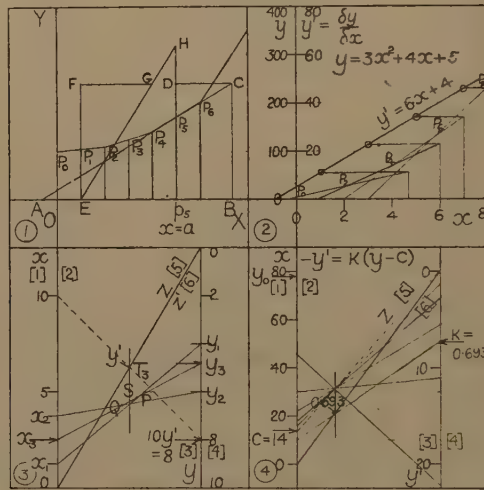
Scales of a nomogram are indicated by square brackets enclosing a scale number, or scale letter. For example, [3] means "scale 3" and [x] means "scale x."

Any graduation figure (number or letter) is placed in front of the scale sign. For example, a reading a [1] means "a graduation mark figured a on scale 1" and is read " a on 1."

a [1] : b [2] means "the join of a on 1 and b on 2." If this join passes through a third point C or c [3], this is indicated by a [1] : b [2] \rightarrow (C or c [3]).

A secor is an ungraduated and unfigured straight line. A secor is used to establish a relation between the points in alignment with it.

Fig. 1.



Transference of a reading a on one scale [2] to another scale [3] is denoted by a [2] \nearrow a [3], which reads "transfer a from 2 to 3."

A normal to a scale or scales passes through opposite or corresponding points. This is symbolized by a [1] \equiv b [2], which reads " a on 1 is opposite b on 2."

The usual notation of the calculus is used but y' and y'' relate to small but not necessarily indefinitely small increments of x and y .

"(AB : CD) \rightarrow E" means "AB intersects CD at E."

2. Differentiation.

1. DERIVED CURVES.

Let the curve $y=f(x)$ be given as an x, y plot to rectangular axes.

Construction 1. (1.1.) To construct an ordinate at $x=a$ equal to $\delta y/\delta x$ for $x=a$.

Let $P_0P_2, P_2P_4, P_4P_6, \dots$, etc. be chords of the curve $y=f(x)$.

Let P_1, P_3, P_5, \dots , etc. be the mid-points of P_0P_2, P_2P_4, P_4P_6 , etc.

Let the chord P_4P_6 be extended indefinitely in both directions meeting the axis OX at A . Make AB equal to any convenient and suitable length (for example 10 units). Let this length be h units. Draw BC parallel to OY to meet $A P_4P_6$ at C . Draw CD parallel to OX to meet the ordinate p_5P_5 (the ordinate at $x=a$) in D .

Then, at $x=a$, $\delta y/\delta x = BC/h$ or p_5D to the scale of OY with its scale numbers diminished h times.

Construction 2. (1.1.) To construct an ordinate at $x=a$ equal to p/q of an ordinate at $x=b$.

For some purposes it may be desired to construct an ordinate p_5H at p_5 of length p/q times BC .

Make p_5D equal to BC . Mark a length p_5E along OX equal to p , make ordinate EF equal to BC , make FG equal to q and parallel to OX , draw EGH to meet p_5D in H . Then p_5H is the ordinate required.

In 1.1, $p=4, q=3$, and $p_5H = BC(4/3)$.

Construction 3. (1.2.) The curve $y=f(x)$ being given, to construct the first derived curve. (y', x).

The curve $y=f(x)$ is plotted to rectangular axes. Draw the figure $P_0P_2P_4P_6 \dots P_n$. Find successive ordinates $y' = \frac{\delta y}{\delta x}$ of the first derived curve by constructions 1 and 2. (1.1.)

The method here given (1.2) is similar to the usual one given in the text-books for constructing a derived curve or a Cartesian plot x, y' . The method may be unsatisfactory in the neighbourhood of a turning point, where y' passes through the value zero. In this case the curve should be broken up into short portions on either side of the turning point and the average value of y' found for each portion. Where the curvature of the curve is smaller the method gives satisfactory results for a moderate range of δx . The method of Construction 4 is, however, more accurate for the same range of δx .

Evidently, the successive derived curves may be obtained in a similar way, the n th derived curve being obtained from the $(n-1)$ th.

If a curve (for example, a seismogram or a cardiogram) is too small to give reasonable accuracy by this method the curve may be enlarged, using a pantograph, microscope, or by geometrical or optical projection.

2. NOMOGRAPHIC DIFFERENTIATION. (1.3 and 1.4.)

If x and $y=f(x)$ be plotted on parallel axes, $\delta y/\delta x$ or y' may be obtained on a segmentary scale connecting the two axes.

1.3 is a general nomogram for finding y' . If corresponding values of x and y are given for consecutive equidistant values of x , y' may be found using nomogram 1.3.

[1] is a uniform scale for x .

[3] is a uniform scale for $y=f(x)$.

[5] is a segmentary scale for $z=y/x$. This scale is convenient but not essential.

Construction 4 (1.3) $x_1, y_1; x_2, y_2; x_3, y_3$ are corresponding values of x and y , x_3 being $\frac{1}{2}(x_1+x_2)$. It is required to find y' for $x=x_3$.

Draw secors $x_1:y_1; x_2:y_2; x_3:y_3$. Let $x_3:y_3$ meet $x_1:y_1$ at P and $x_2:y_2$ at Q.

S is the mid-point of PQ.

ST is parallel to the x and y axes, and meets [5] at T_3 . T_3 is the point y' [5] for $x=x_3$.

A careful drawing-board construction will usually give y' correct to 1 per cent.

This construction is readily applied to pairs of observations and gives the value of y' for any value of x or y .

The segmentary scale is not essential. Since $y=xz$, $1 [1]: y' [5] \rightarrow y' [3]$. Any convenient value of x may be chosen, for example $x=1$ or 10.

In 1.3, if $x=10$ (dotted secor), $y'=y/10$ for $x_3=3$.

The construction may be repeated and y' may be found for a series of consecutive values of x or y . The data so obtained may be used to ascertain whether x and y follow a law such as $y'=-k(x-c)$ or $y'=-k(y-c)$. 1.4 shows the nomographic solution of a differential equation of this type and gives all numerical solutions.

Nomogram 1.4. 1.4 relates to what might be time and temperature observations of a body cooling in a stream of air at constant temperature.

x mins.	0	1	2	3	4	5	6
y °C.	78	46	30	22	18	16	15
y' °/min.		22.2	11.1	5.55	2.77	1.39	

y °C. is the temperature of the body at time x mins. The values of y' may be found by Construction 4. Nomogram 1.4 shows that $-y'=k(y-c)=0.693(y-14)$. Hence, the constant temperature of the stream of air is 14° C. [4] may be graduated for x .

Notes.—1. It may happen, in carrying out Construction 4 (1.3), that the scale zeros are inaccessible. In this case, mark zeros in any convenient positions on [2] and [3], the line joining these zeros being the support for [5] and [6]. Graduate [2] in the same unit as [1], and [3] as [4]. Use [2] and [3] to graduate [5] and [6].

2. Occasionally, the points P and Q (1.3) may lie close together. In this case, $x_1, y_1; x_2, y_2$ must be chosen nearer to x_3, y_3 .

3. Nomographic Solution of Differential Equations.

$-y'=k(y-c)$. (2.0, 3.1, 3.2.)

2.0 is a nomogram in which corresponding values of x and y are joined by straight lines. y' is found from the intersections of these lines, as illustrated in Construction 4. Having found y' , k and c may be found as in 1.4.

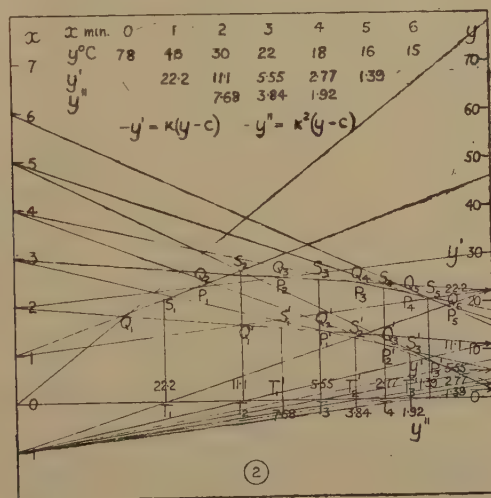
3.1 illustrates an alternative method of finding k and c after y' has been found for two or more values of y .

3.2 indicates a nomographic method of finding corresponding values of x , θ , y' , y after c has been found. A scale for y may be added if desired— $y=\theta+14$.

Thus, all numerical solutions of the differential equation $-y'=k(y-c)$ or of the equivalent equation $y=y_0 \cdot e^{-kx}$ may be obtained graphically by means of the constructions given.

2.0 also indicates how y' may be projected from the segmentary scale on to the y scale, using the point $-1[x]$ as a centre of projection. After this has been done, y'' may be found from x and y' in the same way that y' is found from x and y . In this way the successive derivatives y' , y'' , etc. may be found and projected along the axis of y . The scale for y' , y'' , etc. is easily increased by choosing -2 , -5 , or -10 on x as the centre of projection instead of -1 .

Fig. 2.



$$y' = 2ax + b \quad \text{or} \quad y' = k(x - c).$$

If corresponding values of x and y are given, y' may be found graphically (e. g. by Construction 4) or by taking differences of x and y .

Nomogram 3.3 is a nomogram for x , y' . Since $y'' = 2a$, a may be found from the nomogram by finding y'' (as in 2.0). Also $y' = b$ when $x = 0$. Thus the nomogram gives a and b .

3.1 and 3.4 show the relation between y'' and y' (see fig. 2).

4. Integration.

1. INTEGRAL CURVES.

Construction 5. (4.1.) Given the curve $y=f(x)$ to construct an ordinate at $x=a$ equal to $\sum_{x=0}^{x=10} (y \cdot \delta x)$ or $D^{-1} \cdot y$.

Draw the figure $P_0P_2P_4P_6P_8P_{10}$, the points P having abscissæ $x=0$, $x=2$, $x=4$, etc.

Fig. 3.

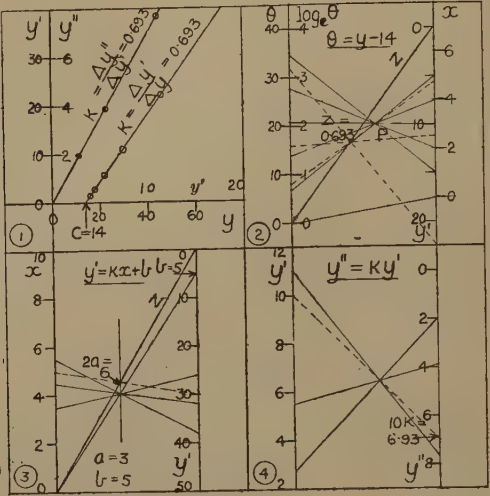
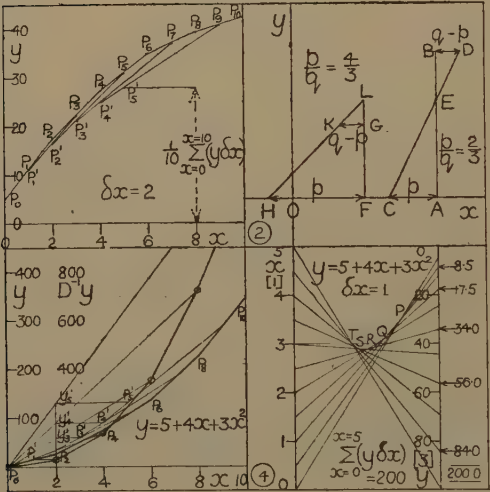


Fig. 4.



Let P_1, P_3, P_5 , etc. be the mid-points of P_0P_2, P_2P_4, P_4P_6 , etc.
 P'_1 coincides with P_1 .
 P'_2 divides P_1P_3 in the ratio 1 : 1.
 P'_3 divides P_2P_5 in the ratio 1 : 2.
 P'_4 divides P_3P_7 in the ratio 1 : 3.
 P'_5 divides P_4P_9 in the ratio 1 : 4,

Let y_1, y_1' be the ordinates of P_1, P_1' ;
 y_2, y_2' be the ordinates of P_2, P_2' , etc.

Hence,

$$\begin{aligned} y_1 + y_3 &= 2y_2', \\ 2y_2' + y_5 &= 3y_3', \\ 3y_3' + y_7 &= 4y_4', \\ 4y_4' + y_9 &= 5y_5'. \\ y_1 + y_3 + y_5 + y_7 + y_9 &= 5y_5'. \\ (y_1 + y_3 + y_5 + y_7 + y_9)\delta x &= 5y_5' \cdot \delta x. \end{aligned}$$

Hence, if an ordinate of length $5y_5'$ be drawn at $x=10$, twice this ordinate is equal to $\sum_{x=0}^{x=10} (y \cdot \delta x)$. If the scale numbers along the axis of y are doubled, the ordinate is equal to $\sum_{x=0}^{x=10} (y \cdot \delta x)$ to the new scale.

To multiply y_5' by 5, y_4' by 4, y_3' by 3, etc., erect an ordinate at $x=2$ equal to y_5' or y_4' or y_3' , etc. (4.3.) Let these ordinates be AB. (2.2.) Join OB and produce OB to meet the ordinates at $x=4, x=6, x=8$, and $x=10$ at C, D, E, F, respectively. These points are marked with circles in 4.3 but they are not lettered. Then the ordinates of C, D, E, and F are 2, 3, 4, and 5 times AB. (See later Construction 7, 4.3.)

If any of these ordinates are too large or too small, the ordinate of A may be altered in length in any desired ratio, using Construction 2, and the scale numbers along the axis of y changed appropriately.

If an ordinate of length $\sum_{x=a}^{x=b} (y \cdot \delta x)$ is required, Construction 5 is applicable, but the numbering 0 to 10 along the x -axis extends from a to b .

The accuracy of the construction is increased by the use of 10 mid-ordinates instead of 5. In this case, $\sum_{x=a}^{x=b} (y \cdot \delta x) = 10y_5' \cdot \delta x$, where $\delta x = (b-a)/10$.

Construction 6. (4.2.) To change the length of an ordinate in the ratio p/q .

Case 1. $q > p$, ordinate AB, $x=OA$.

Make $AC=p$ and $BD=q-p$, AC and BD being parallel to OA and on opposite sides of AB.

CD meets AB at E. Then, $AE=p/q(AB)$.

Case 2. $q < p$, ordinate FG, $x=OF$.

Make $FH=p$ and $GK=p-q$, FH and GK being parallel to OF and on the same side of FG. $(HK : FG) \rightarrow L$.

Then $FL=p/q(FG)$.

Construction 7. (4.3.) To construct the integral curve of $y=5+4x+3x^2$ or $D^{-1}y, x$.

The integral curve is obtained by Construction 5. The ordinates y_1', y_2', y_3', y_4' , and y_5' have been multiplied by 1, 2, 3, 4, 5 respectively

by projection from the origin. The scale numbers along the axis of y have been multiplied by 2 in order to obtain the scale for $D^{-1}y$.

Construction 8. (4.4). Given corresponding values of x and y , to find $\sum_{x=a}^{x=b} (y \cdot \delta x)$.

4.4 shows a graphical method of finding $\sum_{x=0}^{x=5} (y \cdot \delta x)$, where $y=5+4x+3x^2$. Secors connecting corresponding successive values of x and y intersect at P, Q, R, etc. and envelope a curve.

Draw secors $0.5 [1] : P \rightarrow {}_0y_1$, $1.5 [1] : Q \rightarrow {}_1y_2$, $2.5 [1] : R \rightarrow {}_2y_3$, etc. These are the values of y corresponding to the mid-ordinates at $x=0.5$,

1.5 , etc. $\sum_{x=0}^{x=5} (y \cdot \delta x) = {}_0y_1 + {}_1y_2 + {}_2y_3 + {}_3y_4 + {}_4y_5$, since $\delta x=1$.

If the given values of y do not correspond to $\delta x=1$, each mid-ordinate must be multiplied by δx . For example, if $\delta x=h$,

$$\sum_{x=0}^{x=5} (y \cdot \delta x) = ({}_0y_1 + {}_1y_2 + {}_2y_3 + {}_3y_4 + {}_4y_5)h.$$

5. Determination of Laws.

In order to find y in terms of x , it is usual to endeavour to find two functions of y and x , namely $f(y)$ and $F(x)$, which are linearly related. In this case $f(y)=k \cdot F(x)+c$, where k and c are constants. Any suspected linear relation between $f(y)$ and $F(x)$ may be tested nomographically. If secors connecting corresponding values of $f(y)$ and $F(x)$ are concurrent, the linear relation holds, and the values of k and c may be found from the point of concurrency. If, owing to errors of observation, the secors are not quite concurrent, visual examination of their intersections enables the position of the point of concurrency to be estimated, and thus enables the best values of k and c to be found.

$$1. \quad y=a+b \cdot x+c \cdot x^2 \quad (5.1).$$

In this case, $\Delta y/\Delta x=b+2c \cdot x$; $\Delta(\Delta y)/(\Delta x)^2=2c \cdot \Delta x$.

$$x : \Delta y/\Delta x \rightarrow 2c [z].$$

Let the following values of x and y be given :—

x	0	10	20	30	40	50
y	4	1.2	2.6	4.6	7.2	10.4
$\Delta y/\Delta x$..		.08	.14	.20	.26	.32
$\Delta^2 y/(\Delta x)^2$.006	.006	.006	.006

$\Delta y/\Delta x$ may be found graphically or by calculation.

5.1 shows that the secors intersect at a point given by

$$x : \Delta y/\Delta x \rightarrow 0.006 [z]. \quad \therefore 2c=0.006, c=0.003.$$

Also, when $x=0$, $\Delta y/\Delta x=0.05$ and $y=0.4$.

$$\therefore b=0.05 \quad \text{and} \quad a=0.4.$$

Hence

$$y=0.4+0.05x+0.003x^2.$$

Experimental observations must be smoothed before applying a nomographic test.

$$2. \ y = a \cdot x^b \quad (5.2).$$

$\log y = b \cdot \log x + \log a$, hence $\log y : \log x \rightarrow b [z]$ and $\log y = \log a$ if $\log x = 0$.

The following data relate to the speed v , knots of a steamer of I.H.P.
P:—

v	10	12	14	16	18	20
P	1066	1912	3216	4951	7361	10360

The secors $\log v : \log P$ are concurrent.

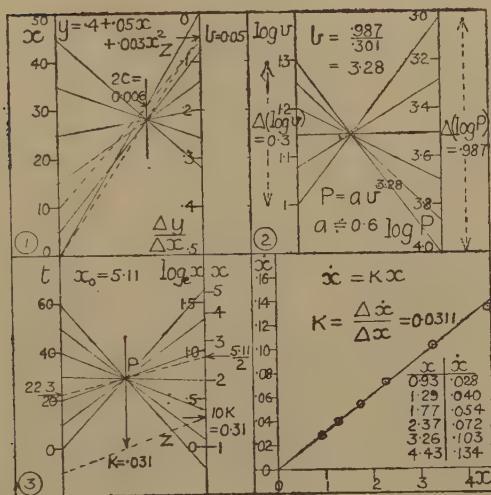
$$\Delta(\log P) / \Delta(\log v) = 3.28 = b.$$

$$P = 1066 \text{ when } v = 10 \text{ gives } a = 0.6.$$

Hence,

$$P = 0.6v^{3.28}.$$

Fig. 5.



$$3. \ y = a \cdot e^{bx} \quad (\text{Fig. 6})$$

$$\log_e y = b \cdot x + \log_e a$$

Hence, $\log_e y : x \rightarrow b [z]$ and $\log_e a = \log_e y$ when $x = 0$.

Fig. 6 illustrates how to find the velocity constant k for the conversion of dibromosuccinic acid into bromomaleic acid by boiling with water. The data are the experimental observations of J. H. van't Hoff, given in 'Chemical Statics and Dynamics,' by J. W. Mellor (1904).

\dot{x} is found as in fig. 2.

Secors $t : \dot{x}$ are concurrent at P or 0.031 [z].

The velocity constant $k = 0.031$.

5.3 shows that secors $t : \log_e x$ are concurrent at P or 0.031 [z]. A scale for x is easily marked adjacent to $\log_e x$. The time for transformation of half of the initial quantity of the reacting substance is easily obtained—the dotted secor gives 22.3 minutes.

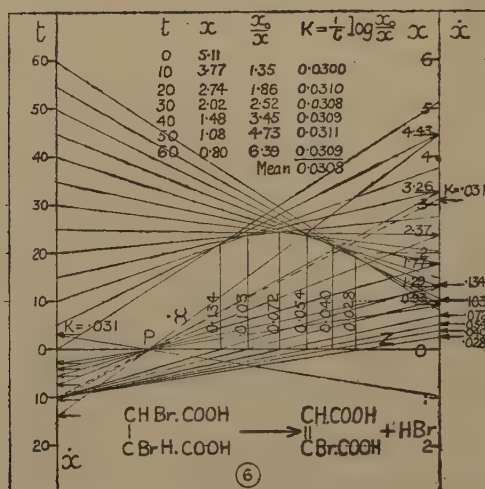
The nomogram provides a test for a first-order reaction.

5.4 shows that $\dot{x}=0.031x$ and gives the velocity of the reaction for any value of x .

ADDENDUM.

Dr. C. J. Smith, *Phil. Mag.* vol. xxxiii. p. 775, states "In a recent paper it was stated 'It is impossible to obtain rates of cooling with sufficient accuracy by drawing tangents to a cooling curve'" (Hansel, *Proc. Phys. Soc.* liv. p. 163). Dr. Smith then describes the well-known mirror method of drawing a normal to a curve as though in ignorance of this method I had made a rash and unjustifiable statement. I, too, have used this method (first communicated to me by Prof. Ferguson), and I find that the accuracy is improved by using a set-square with its right angle against the mirror. Even so, errors as great as 5 per

Fig. 6



cent. are common, and larger errors still unless great care and repetition of work is exercised. Moreover, such errors are to be expected when a rotation of the mirror by 0.1° produces moderate error in the gradient.

My actual remarks were made in relation to examination questions and crude (indeed incorrect) interpretations of Newton's law of cooling.

Let the gradient of $y=\log x$ be found, by this method, over a range $x=3$ to 9.

If the gradient be plotted against $1/x$ it becomes apparent that such results are quite inadequate to decide what is the true relation between the gradient and $1/x$.

I therefore affirm the utter futility of the method of drawing tangents for students of School Certificate experience. I may add that those experienced in graphical methods obtain precarious results much too uncertain to decide with any degree of confidence between a linear law and alternative possibilities.

LXVI. *Roots of Sin $z=z$.*

By A. P. HILLMAN and H. E. SALZER *

(Mathematical Tables Project, National Bureau of Standards, U.S.A.).

[Received March 15, 1943.]

THE following table gives the first ten non-zero roots of $\sin z=z$ in the first quadrant to six decimal places. Obviously the roots are symmetrically situated in the four quadrants.

Equating real and imaginary parts of $\sin (x+iy)=x+iy$, one obtains

$$\sin x \cosh y = x \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$\cos x \sinh y = y \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Equivalent to (1) and (2) are equations

$$x = \coth y (\sinh^2 y - y^2)^{\frac{1}{2}} \equiv x_1(y), \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$x = \arccos (y/\sinh y) \equiv x_2(y). \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Roots of $\sin z=z$ where $z=x+iy$.

<i>n.</i>	<i>x.</i>	<i>y.</i>
1.....	7.497676	2.768678
2.....	13.899960	3.352210
3.....	20.238518	3.716768
4.....	26.554547	3.983142
5.....	32.859741	4.193251
6.....	39.158817	4.366795
7.....	45.454071	4.514640
8.....	51.746768	4.643428
9.....	58.037662	4.757515
10.....	64.327234	4.859917

The problem of solving $\sin z=z$ is therefore reduced to the finding of the real zeros of the real function $x_1(y)-x_2(y)$. The imaginary part of the n -th root was obtained as follows:—Three values of x_1-x_2 (at interval in y not more than 0.01) were calculated (for the appropriate branch of x_2) in the neighbourhood of G. H. Hardy's approximation

$$y = \log (4n+1)\pi \dagger.$$

Quadratic inverse interpolation was then used to find the tabulated y . $x_2(y)$ was taken as the tabulated x . Finally

$$\Delta y = (x_1 - x_2) \bigg/ \frac{d(x_1 - x_2)}{dy} \quad \text{and} \quad \Delta x = \Delta y \frac{dx_2}{dy}$$

were calculated as first approximations to the errors in y and x respectively. Δy and Δx were always less than 10^{-7} .

* Communicated by the Authors.

† G. H. Hardy, *Mess. of Math.*, n. s., vol. xxxi. pp. 161-165 (Mar. 1902).

LXVII. *Notices respecting New Books.*

Elementary Physical Chemistry. By MERLE RANDALL and LEONA YOUNG. [Pp. xxxii+455.] (Randall & Sons, Berkeley, California, U.S.A. Price \$4.50.)

THIS unorthodox book offers an intensive one-year course in modernized classical physical chemistry. Restricted largely to equilibrium theory, it omits colloids and surface phenomena, structure-sensitive properties of liquids, and modern structural theory. The treatment of states of aggregation is inadequate and the chapter on "systems not at equilibrium" is unworthy of the remainder of the book. The authors make little use of molecular and kinetic theories and prefer a formal treatment.

The arrangement of subject-matter is unusual. A number of topics familiar to all students from elementary laboratory work are successively subjected to thorough discussion. Gradually all the important physical concepts are introduced in an attractive way. Other topics are then presented, the principles elaborated and used to explain well-known laboratory methods and industrial processes. Principles and practice are closely integrated. Thermodynamic equations are introduced without derivation, their meaning and limits of applicability being defined with great care. The persistent emphasis on the quantitative approach is most commendable. Accurate data are provided as an integral part of the book in the form of numerous tables and graphs. Seven hundred numerical exercises provide practice in computation. A nice sense of balance is preserved in reference to classical and modern literature.

It is to be hoped that the photostat reproduction of typescript notes does not secure wide adoption in book production. This text is extremely tiring to the eye, and the limited range of type does not provide adequate differentiation of text proper, exercises and legends attached to diagrams.

This book should be of great value to teachers of classes of students with divergent interests, where lack of time frequently precludes any formal study of thermodynamics. A keen appreciation of common difficulties is combined with many novel features in presentation.

W. WILD.

Statistical Tables for Biological, Agricultural and Medical Research. By R. A. FISHER and F. YATES. Second Edition. [Pp. viii+98.] (Oliver and Boyd, Edinburgh, 1943. Price 13s. 6d.)

THESE tables—first published in 1938—are of such importance for modern statistical practice and cover so wide a field of applications that a welcome to the new edition may usefully be accompanied by a brief indication of the changes made. Four new tables are included without alteration of the numbering of the other tables; two of these deal with the Fisher-Behrens test, one with the dilution method in bacteriology. The fourth (Table VIII., I.) is based on W. L. Stevens' arrangement of fiducial limits for binomial and Poisson distributions, and it might be advisable for users of this table to add to it a cross-reference to the addendum on p. 25. Table XVII., "balanced incomplete blocks," is considerably altered by changes in presentation due to use of cyclic substitutions and by additional solutions; there are corresponding changes and some extensions in the Introduction. Errata found in the first edition have been corrected and a list of the few numerical ones is given.

G. S.

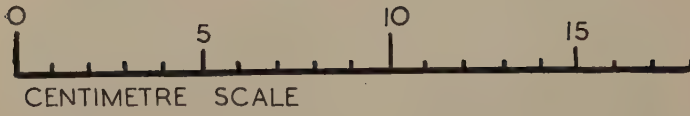
[The Editors do not hold themselves responsible for the views expressed by their correspondents.]



Half actual size.

Hot mercury arc source, $\lambda 5461$.
Interferometer gap, 1.5 mm.
Source-screen distance, 2.75 metres.

CENTIMETRE
SCALE



A.



Sharp single fringe.

Source-screen distance, 1.75 metres.
Interferometer gap, 1.5 mm., λ 5461.

B.



Hyperfine structure.

Source-screen distance, 2.75 metres.

C.



Source-screen distance, 4.5 metres.

D.



100th
ring

Source-screen distance, 1.75 metres.

1st
ring

A.



Two holes source.

..

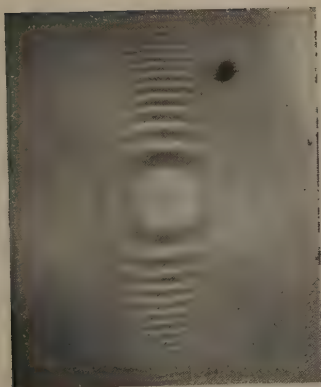
B.



Horizontal slit source.

—

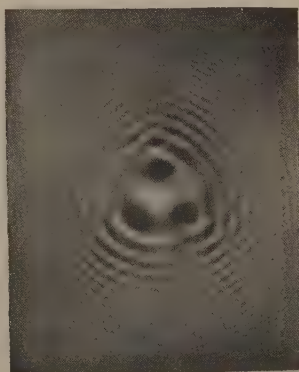
C.



Crossed slit source.

+

D.



Y-shaped source.

Y

LXVIII. *Anomalous Viscosity shown in Oil Flow through Engine Bearings.*

By S. M. NEALE, D.Sc.*

[Received February 10, 1943.]

SUMMARY.

In the following paper it is shown that the usual assumption of proportionality between shearing stress and velocity gradient is incapable of explaining recent data on oil flow through engine bearings.

It is suggested that at high rates of shear lubricating oils undergo molecular orientation, causing a large fall in viscosity, and that the accepted theories of fluid lubrication are inadequate.

Introduction.

IN his classical analysis⁽¹⁾ of fluid flow in a bearing, Reynolds pointed out that his assumption of a constant viscosity coefficient, that is to say of direct proportionality between shear stress and velocity gradient, had no foundation other than that afforded by experimental fact. Rigid confirmation of this assumption is afforded only by measurements at the relatively low rates of shear encountered in viscometers of the Couette, capillary, or other conventional types.

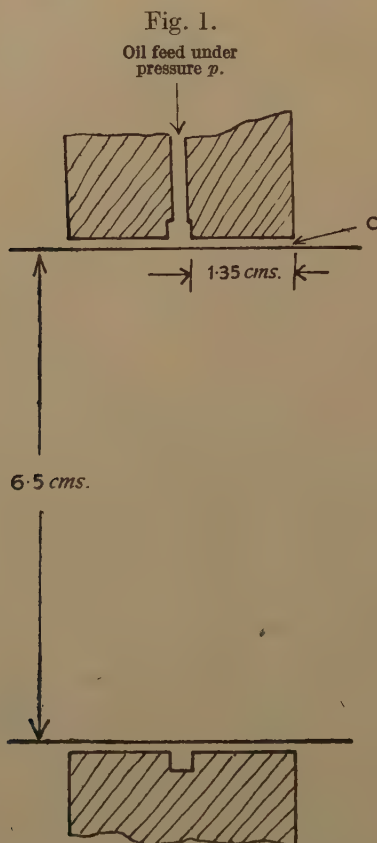
It is now well known that in solutions containing chain molecules of great length, such as cellulose and its derivatives, rubber, myosin, and many synthetic high polymers, the viscosity coefficient is not a constant, but decreases steadily with increasing rate of shear^{(2) (3)}. For example, it was shown by Farrow, Lowe and Neale⁽⁴⁾ that solutions of starch in aqueous formaldehyde showed such an effect over a range of velocity gradient from 10^{-2} to $10^{3.5}$ sec.⁻¹, whereas castor oil over the range 10^{-2} to $10^{2.5}$ sec.⁻¹ showed a constant viscosity coefficient.

It is generally considered that this drop in apparent viscosity is due to molecular orientation. If this is so, it would be anticipated that molecules of less pronounced anisometric shape, such as those of lubricating oils, would show such an effect at higher rates of shear, when the Brownian movement may be insufficient to maintain their normal random orientation. Bucheim, Stuart and Menz⁽⁵⁾ claim to have shown, by means of extremely sensitive refractive index measurements, that even benzene undergoes very slight molecular orientation at velocity gradients of the order 10^3 sec.⁻¹. Kyropoulos⁽⁶⁾ attempted to show that certain viscous lubricating oils show a fall in viscosity coefficient at rates of shear of the order 10^4 sec.⁻¹. His rate of shear sufficed to produce a viscosity drop of the order of 20 per cent. only, and the design of his apparatus was such that the necessary corrections for temperature rise

* Communicated by the Author

were larger than the observed effect. The author, in 1937, referred briefly⁽⁷⁾ to comparatively rough experiments, with rates of shear up to 6×10^4 , which indicated a drop of about 70 per cent. in the viscosity coefficient of castor oil. The work of Kyropoulos and the writer does not appear to have influenced the opinions of engineers, who generally still assume that the viscosity coefficient of an oil is a function of temperature and pressure only.

This "classical" view has been supported by the experiments of McKee and McKee⁽⁸⁾, and of Kingsbury⁽⁹⁾, who worked up to velocity



gradients of the order 10^5 . Kingsbury's paper is notable in that it included a thorough analysis of the temperature gradients within the oil film itself. These workers, however, used oils of relatively low viscosity (about 40 centipoise at 77° and 100° F. respectively), in which shear orientation sufficient to cause a substantial fall in viscosity might require rates of shear higher than they employed.

Spiers' Experiments.

In the present communication use is made of the recent data of Spiers⁽¹⁰⁾, as evidence of anomalous viscosity of oil at high rates of shear. Spiers

has given an account of careful measurements of the flow of lubricating oil under feed pressures ranging between 40 and 100 lbs./in.², through the bearings of an experimental automobile engine. Such bearings are characterized by fine clearances and high rotational speeds, and a relatively viscous lubricant is usually employed. In them rates of shear up to 10^6 sec.^{-1} are encountered.

Now such a bearing (fig. 1) may be regarded as a viscometer, in which the apparent viscosity of the fluid undergoing rapid shear owing to the rotation of the shaft can be calculated from the known dimensions and the rate of axial flow of oil under the known pressures. Proceeding on this basis it will be shown that the increase of rate of flow with increasing rotational speed indicates that the apparent viscosity of the oil decreases under shear to a small fraction of its normal value. Furthermore, in Spiers' experiments the flow of oil through the bearing is so rapid that this effect cannot be wholly attributed to rise in oil temperature.

The calculations which follow refer only to the "main" bearings tested by Spiers (see his Table II.) with various clearances and to the oil "B" which, at the feed temperature of 70° , had a viscosity of 0.32 poise, measured in a viscometer of conventional type. The results of the flow measurements for "big-end" bearings have also been analysed and found to agree with the main conclusions of this paper. Spiers' data for the other less viscous oil "C" are insufficient.

Notation (in C.G.S. units).

a = Radius of shaft (= 3.25 cm.).

b = Internal radius of bearing.

l = Effective half bearing length = length of axial flow path (= 1.35 cm.).

c = Radial clearance = $b - a$.

N = Speed of rotation revs./sec.

V = Flow of oil c.c./sec.

η = Viscosity coefficient at low velocity gradient.

$\bar{\eta}$ = Mean viscosity coefficient at velocity gradient du/dx or dv/dx .
(Mean shear stress/velocity gradient.)

ρ = Density of oil (= 0.84 at 70°C.).

h = Specific heat of oil (= 0.5 approx.).

θ = Mean temperature of oil.

p = Pressure of oil feed.

f = Mean shear stress due to rotation.

du/dx = Velocity gradient for circumferential shear (due to rotation).

dv/dx = Velocity gradient for axial shear due to oil flow under pressure p .

dv/dx = Resultant velocity gradient.

The rate of laminar flow (parallel to the axis) of a viscous fluid through an annulus is given by Lamb's expression⁽¹¹⁾

$$V = \frac{\pi p}{8\eta l} \left\{ b^4 - a^4 - \frac{(b^2 - a^2)^2}{\log_e b/a} \right\} \dots \dots \dots (i)$$

In the present case where b and a are nearly equal, this reduces to *

$$\frac{\pi p a c^3}{6 l \eta},$$

which is identical with the equation for viscous flow through a plane slit without ends. As, in Spiers' bearing (fig. 1), oil flows through two such annuli in parallel,

$$V = \frac{\pi p a c^3}{3 l \eta}, \quad \dots \dots \dots (ii)$$

So long as the flow remains laminar in character and the viscosity-coefficient remains constant, V should not be affected by simple rotation of the inner wall of the annulus. It has been shown by Cornish⁽¹²⁾ and by Fage⁽¹³⁾ that the flow of water through such an annulus is independent of speed of rotation up to a certain critical speed, above which the axial flow decreases. Equation (ii) may therefore be used to obtain maximum values of $\bar{\eta}$ from Spiers' data.

Correlation of $\bar{\eta}$ with rotational velocity gradient du/dx alone.

$$du/dx = 2\pi a N/c \quad \dots \dots \dots (iii)$$

and

$$f = \bar{\eta} \cdot du/dx, \quad \dots \dots \dots (iv)$$

whence the power absorbed in shearing the oil film, assuming the shaft to be centralized in the bearing, is

$$16\pi^3 a^3 l N^2 \bar{\eta} / c, \quad \dots \dots \dots (v)$$

and the corresponding theoretical temperature rise, which, added to the oil inlet temperature (70° C.), gives the *mean* temperature of the oil leaving the annulus, is given by

$$\theta - 70 = \frac{16\pi^3 a^3 l N^2 \bar{\eta}}{4.18 \times 10^7 c V \rho h} \quad \dots \dots \dots (vi)$$

The application of equations (ii) to (vi) to Spiers' data gives the values tabulated on p. 581. All quantities are expressed in absolute C.G.S. units. du/dx and f are calculated from the circumferential shear only. c , N and V/p are the data taken from Spiers' paper.

In fig. 2 the values of $\bar{\eta}$, uncorrected for temperature, are plotted against the logarithms of the corresponding velocity gradients. It will at once be seen that the viscosity coefficient, far from being independent of the velocity gradient, as is found for lubricating oils in normal viscometry, falls very markedly indeed at high rates of shear. The crux of the matter is, of course, that the amount of axial flow is far greater than that predicted by the classical formulation, and it is now suggested that this is due to an orientation of the oil molecules under shear.

Such orientation will presumably be parallel to the direction of rotation of the shaft, whereas the measured flow takes place parallel to

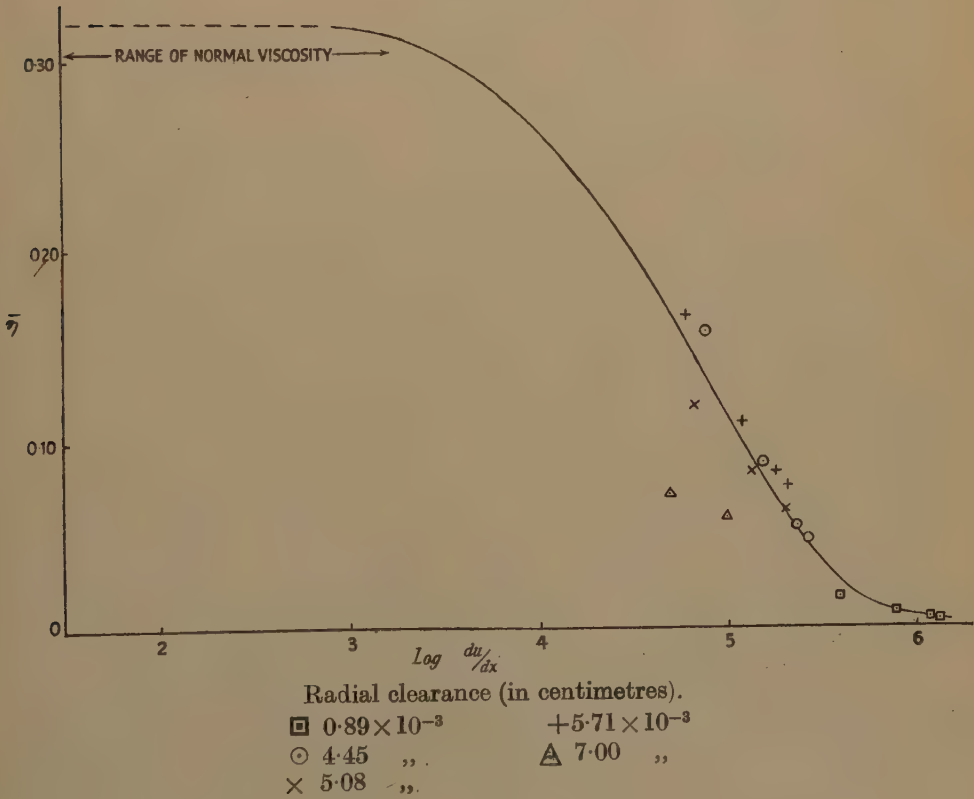
* Spiers has, in his paper, reduced equation (i) incorrectly, apparently through an insufficient expansion of $\log_e b/a$.

TABLE I.

$c=0.89 \times 10^{-3}$ cm. (approx.).					$c=4.45 \times 10^{-3}$ cm.			
N	16.7	33.3	50	58.3	16.7	33.3	50	58.3
du/dx	3.83	7.62	11.5	13.1	0.77	1.53	2.30	2.68×10^5
V/p	0.109	0.196	0.323	0.425	1.41	2.52	4.08	5.29×10^{-6}
$\bar{\eta}$	1.62	0.90	0.54	0.41	15.7	8.8	5.4	4.7×10^{-2}
f	6.2	6.86	6.26	5.44	12.1	13.5	12.5	12.7×10^3
θ	82.4	83.8	82.6	81.0	71.8	72.4	72.0	71.8

$c=5.08 \times 10^{-3}$				$c=5.71 \times 10^{-3}$				$c=7 \times 10^{-3}$	
N	16.7	33.3	50	16.7	33.3	50	58.3	16.7	33.3
du/dx	0.66	1.32	2.0	0.60	1.19	1.79	2.08	0.485	0.97×10^5
V/p	2.81	3.99	5.26	2.81	4.28	5.63	6.15	12.2	14.8×10^{-6}
$\bar{\eta}$	11.8	8.3	6.3	16.6	10.9	8.3	7.6	7.1	5.9×10^{-2}
f	7.76	10.9	12.6	10.0	13.0	14.9	15.9	3.46	5.68×10^3
θ	70.6	71.2	71.6	70.8	71.4	71.8	72.0	70.0	70.2

Fig. 2.



the axis of the shaft; that is to say, at right-angles to the direction of molecular orientation. It is quite likely that in the supposed oriented state the viscosity coefficient will vary with the direction of shear. As the torque on the bearing is not known this cannot be ascertained from the experiments of Spiers. It is, however, legitimate to suppose that molecular orientation will have some effect upon the viscosity coefficient for all directions of shear. At the high shear stresses set up in the oil in these experiments, it is conceivable that a straightening of the molecules, by rotation about the single valency bonds, may take place, in addition to their orientation.

Correlation of $\bar{\eta}$ with the resultant velocity gradient dv/dx .

It is apparent from fig. 2 that, although the correlation of $\bar{\eta}$ with the rotational velocity gradient alone suffices to demonstrate the existence of a drop in viscosity under shear, it fails to bring into line the two experimental points for the largest clearance. This apparent discrepancy is, however, removed if the velocity gradient of axial flow due to the oil feed pressure p is also taken into account. This velocity gradient is, unlike du/dx , not a constant for given values of c and N , but varies across the annulus, being zero at the centre surface of the annular space and a maximum at the shaft bearing surfaces.

The shear stress causing axial flow is $\frac{px}{l}$, if x be the distance of any point from the central cylindrical lamina of the annular space. If the viscosity coefficient is independent of velocity gradient, the maximum velocity gradient due to axial flow is

$$\frac{dw}{dx} = \frac{pc}{2\eta l} = \frac{3V}{2\pi ac^2} \quad \dots \dots \dots \text{(vii)}$$

(at $x = \frac{c}{2}$), and this is therefore readily derived from the experimental values of V . If, however, as is now suggested, the viscosity is an unknown function of the velocity gradient, the distribution of axial velocity across the annulus is indeterminate. As $\bar{\eta}$ appears to fall with increasing velocity gradient, the maximum value of $\frac{dw}{dx}$ will be greater than that given by equation (vii). In order to make possible an approximate allowance for $\frac{dw}{dx}$, relative to $\frac{du}{dx}$, however, equation (vii) has been used to calculate the approximate values of $\frac{dw}{dx}$ given in Table II., V being taken at the "standard" pressure of 70 lbs./in.². Using them, the resultant velocity gradient values $\frac{dv}{dx}$ are given by

$$\left(\frac{dv}{dx}\right)^2 = \left(\frac{du}{dx}\right)^2 + \left(\frac{dw}{dx}\right)^2 \quad \dots \dots \dots \text{(viii)}$$

TABLE II.
(in C.G.S. units.)

$C=0.89 \times 10^{-3}$ cm.				
N	16.7	33.3	50.0	58.3
V	0.526	0.946	1.56	2.05
$\bar{\eta}$	1.62	0.90	0.54	0.41×10^{-2}
du/dx	3.83	7.62	11.5	13.1×10^5
dw/dx	0.98	1.75	2.9	3.8×10^5
dv/dx	3.96	7.82	11.9	13.6×10^5
$C=4.45 \times 10^{-3}$ cm.				
N	16.7	33.3	50.0	58.3
V	6.8	12.15	19.7	25.5
$\bar{\eta}$	15.7	8.8	5.4	4.7×10^{-2}
du/dx	0.77	1.53	2.30	2.68×10^5
dw/dx	0.50	0.90	1.46	1.89×10^5
dv/dx92	1.78	2.73	3.28×10^5
$C=5.08 \times 10^{-3}$ cm.				
N	16.7	33.3	50.0	
V	13.5	19.2	25.4	
$\bar{\eta}$	11.8	8.3	6.3	$\times 10^{-2}$
du/dx	0.66	1.32	2.0	$\times 10^5$
dw/dx	0.77	1.09	1.44	$\times 10^5$
dv/dx	1.01	1.71	2.46	$\times 10^5$
$C=5.71 \times 10^{-3}$ cm.				
N	16.7	33.3	50.0	58.3
V	13.5	20.6	27.1	29.6
$\bar{\eta}$	16.6	10.9	8.3	7.6×10^{-2}
du/dx	0.60	1.19	1.79	2.08×10^5
dw/dx	0.61	0.93	1.22	1.33×10^5
dv/dx86	1.51	2.16	2.47×10^5
$C=7.0 \times 10^{-3}$				
N	16.7	33.3		
V	58.8	71.3		
$\bar{\eta}$	7.12	5.86	—	$\times 10^{-2}$
du/dx	0.485	0.97	—	$\times 10^5$
dw/dx	1.76	2.14	—	$\times 10^5$
dv/dx	1.83	2.35	—	$\times 10^5$

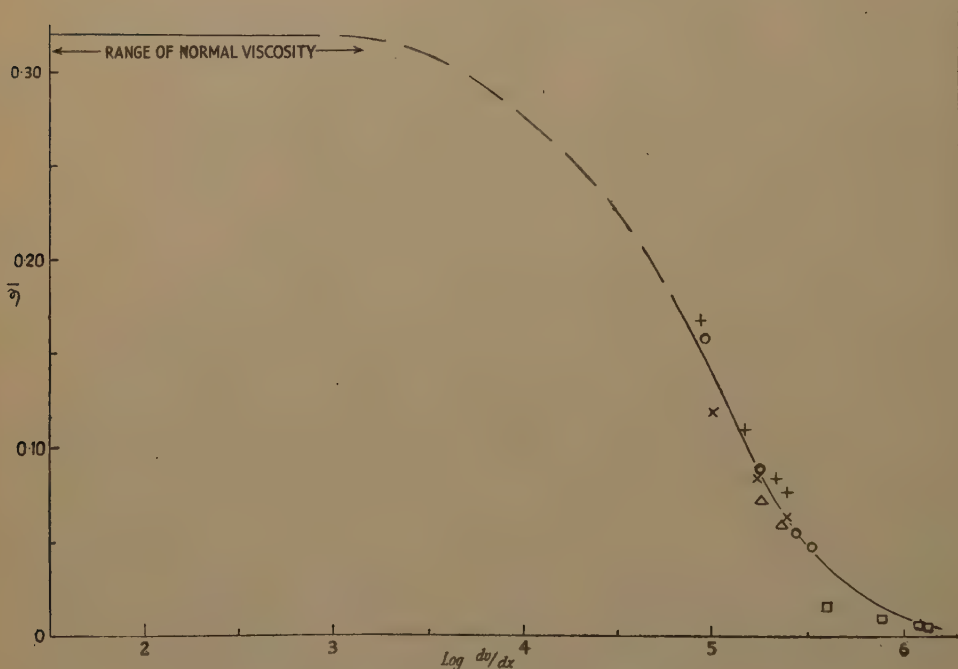
In fig. 3 $\bar{\eta}$ is now plotted against the resultant velocity gradient, and the points for the maximum clearance now lie close to the smooth curve.

Since, however, $\bar{\eta}$ and $\frac{dw}{dx}$ are both linear functions of the measured flow

rate V , this agreement has by no means the same significance as the viscosity drop demonstrated by fig. 2, where $\bar{\eta}$ was calculated from V , and $\frac{du}{dx}$ from the rate of rotation alone. The consideration of the $\frac{dw}{dx}$ values, however, suggests that the physical explanation of the abnormally high values of V for the largest clearance is as follows:—

For a given feed pressure p , the axial shear stress $\frac{pc}{2l}$ increases with c , and when c has reached 7×10^{-3} cm. $\frac{pc}{2l}$ has reached a value sufficient

Fig. 3.



Radial clearance (in centimetres).

□ 0.89×10^{-3}	+ 5.71×10^{-3}
○ 4.45 „	△ 7.00 „
× 5.08 „	

to cause molecular orientation and excessive flow, to an extent comparable with the effect due to rotation.

The points for the smallest clearance ($c=0.89 \times 10^{-3}$ cm.) still lie rather below the curve, but this discrepancy is of little import owing to the uncertainty in the measurement of such a small clearance, and in the absence of any allowance for temperature rise.

The present analysis of the data is only approximate, as is justified by the errors inherent in the experiments, which were not designed for

a specific test of this theory. Nevertheless, it is evident that Spiers' data cannot easily be reconciled with the usual assumption of constant viscosity coefficient at constant temperature, but strongly suggest a substantial fall in viscosity at high velocity gradients.

There are, however, several factors obviously neglected in this simple analysis, and these will now be discussed in turn.

1. *Rise in Oil Temperature.*

This is the explanation put forward by Spiers, but it can easily be shown that an impossibly large quantity of heat would be required to bring about the necessary effect. For instance, when $c=5.7 \times 10^{-3}$ cm. and $N=50$, V is 27.1 c.c. per sec. at the standard pressure of 70 lbs./in.², and $\eta=0.083$ poise. From Spiers' viscosity temperature curve, assuming uniform temperature across the thickness of the oil film, it appears that the oil would have to reach a mean temperature of 115° C., and an exit temperature of about 140° C., to reduce η from 0.32 to 0.083. As the measured oil inlet temperature is 70° C., this would require a heat intake by the lubricating oil in the one bearing alone of 3300 watts, giving about 30 kw. for the ten crankshaft bearings. As the engine under the conditions of test is developing 28 B.H.P., this would mean that more energy is used in heating the oil during its passage from the inlet thermometer to the exit from the clearance space than is given out as useful work! According to Spiers' measurement of thermocouple temperature in the bearing itself, the latter attains a temperature of 90° C. (where $\eta=0.18$) under the conditions detailed above. The oil flowing at 27.1 ccs./sec. could, however, hardly undergo a temperature rise of 20° C. within the clearance space without acquiring a still excessive quantity of energy; the rate would be about 1000 watts per bearing.

The explanation of the fact that the metal surfaces attain a higher temperature than the bulk of the oil between them, lies in a consideration of the temperature gradients within the oil film itself. Kingsbury⁽⁹⁾ was one of the first to point out the importance of this effect. In his experiments, where the oil film did not flow through the annulus, the heat resulting from the shearing of the oil was conducted outwards to the metal surfaces, so that the temperature of the oil film rose to a maximum in the middle and fell off towards the walls. In Spiers' experiments, on the other hand, most of the heat is carried away by the flowing oil itself.

The rate of flow of oil is greatest in the middle of the annulus, and here, therefore, the temperature rise is least. As the walls of the annulus are approached the rate of flow parallel to the axis of the shaft decreases towards zero, and consequently the temperature rises rapidly. On account of the low thermal conductivity of oil, temperature gradients within the film may be very steep, as Kingsbury showed.

To evaluate them in the present instance, and to make allowance for this effect on the flow of oil, would require a knowledge of the viscosity-shear-temperature relations, and a mathematical process too intricate

and laborious to be justifiable, considering the probable errors of experiment.

The effect of allowing for the variation of temperature across the oil film will be to increase the significance of the temperature effect upon the flow of oil, since the temperature rise and the consequent fall in viscosity are greatest near the walls, where the axial shearing stress px/l is a maximum.

Even so, however, it does not seem at all probable that temperature rise alone could account for the observed effect in the instance quoted above.

Considering the data for the smallest clearance used by Spiers, an explanation based upon temperature rise alone seems even less feasible.

Thus, to account for the observed rate of flow when $c=0.89 \times 10^{-3}$ cm. and $N=50$, in terms of viscosity drop due to temperature rise alone, it would be necessary to suppose that the effective mean oil temperature in the clearance space is close to 300°C . It is clear that even though the calculated values of θ in Table I. represent a minimum, temperature rise alone cannot account for the observed effect.

As the shaft and bearing, apart from a thin Babbit metal lining, consist wholly of steel, the clearance may be assumed unaffected by thermal expansion.

2. Eccentricity of the Shaft Position under Load.

With the shaft stationary and the eccentricity equal to the radial clearance, the viscous flow can increase up to 2.5 times the concentric value⁽¹²⁾. At the speeds tested, however, the eccentricity for an assumed load on the bearing of 100 kg. calculated according to the Sommerfeldt equation⁽¹⁴⁾, and using the reduced values of $\bar{\eta}$ given in Table I., is always small. For example, in the case quoted under (1), e/c is only 0.07. The corresponding increase over the concentric value for fluid friction, which led to the values of θ given in Table I., is quite negligible.

3. Disturbance of Laminar Flow in the Oil Film.

If this occurs, then Cornish⁽¹²⁾ and Fage⁽¹³⁾ have shown, from the behaviour of water in such an annulus, that the axial flow should decrease at the higher rates of revolution.

4. Increased Axial Flow due to Pumping Action caused by the Rotating and Reciprocating Load.

This was considered by Spiers and dismissed as negligible, since (a) the V - p relation is a straight line passing through the origin, (b) V is independent of the load on the engine.

It may therefore be concluded that the most plausible explanation of the abnormally high observed values of axial flow lies in a substantial drop in viscosity caused by molecular orientation under shear. In the approximate analysis made above, it has been assumed that the effective viscosities tending to resist circumferential and axial shear are the same, that is to say, that the oriented molecules slip past each other equally

readily in a lengthwise and in a crosswise direction. This may not be strictly true, but a precise analysis of the system must await more accurate determinations of the relation between shearing stress and velocity gradient.

Referring again to Table I., it appears that there is a suggestion that at very high rates of shear the shear-stress becomes independent of the velocity gradient. Should this indeed prove to be the case, a uniform distribution of velocity gradient could no longer be assumed, and abnormally high slip might occur along certain planes.

It might also be inferred that orientation would tend to resist the onset of turbulence, and that the fundamental difference between lubricants and non-lubricants is that the former undergo molecular orientation when sheared in narrow channels, whereas with the latter such an effect is normally prevented by the occurrence of turbulence.

Application of the Theory to Engineering Problems.

The reduction of viscosity due to shear alone will, if fully substantiated, evidently necessitate a revision of the mathematical theory of fluid lubrication. The qualitative conception of a revolving loaded shaft supported by fluid pressures will no doubt be maintained. The calculated frictional coefficient of the bearing at high speeds will be lower than that deduced from classical theory, and it will not be necessary to postulate excessive rise in temperature to explain the well-known fact that the frictional torque does not increase in proportion to the speed of rotation.

The fact that an automobile engine can operate successfully with a lubricant of high viscosity is in itself hardly explicable on the classical theory, since a simple calculation⁽⁷⁾ shows that this theory predicts an excessively large loss of power in oil shear.

It is not yet possible to say how the theoretical load-carrying capacity of a bearing would be affected by the new conception, since the viscous resistance of a fluid in the oriented condition may vary with the direction as well as with the magnitude of shear.

It might further be expected that the viscous resistance to motion would be a minimum when the shear of the lubricant is uniform, as in a full circular bearing, since a finite time would be required to set up the condition of molecular orientation and reduced viscosity. Such a time would probably be of the order 10^{-4} second. Speculation can, however, not be carried too far without further experimental evidence, which it is hoped shortly to provide, of the nature of the effect.

Supplementary Note on Capillary Flow Experiments.

The writer has carried out a series of experiments in which castor oil and Mobiloil BB are expelled through water-cooled steel capillary tubes under driving pressures up to 150 kg./cm.^2 . In these tests velocity gradients at the capillary wall reached $2 \times 10^5 \text{ sec.}^{-1}$ at 20°C . The apparent viscosity fell to about one half its normal value when the maximum rate of flow was reached. However, after correcting for the

effect of hydrostatic pressure on viscosity, and for the estimated variation of temperature across the capillary, a drop of viscosity due to molecular orientation could not be substantiated with certainty. It was, however, found that when a glass capillary was used, or when the high speed jet of oil was examined between crossed Nicol prisms immediately on leaving the orifice, that the boundary layers were strongly double refracting. This may be taken as evidence of molecular distortion or orientation due to the shear. The extinction angle was at about 45° to the streamlines. This indicates that the orientation is only in its initial stages⁽²⁾, and is probably not sufficient to cause a very large fall in the viscosity. The persistence of the birefringence for some distance after the oil had left the capillary showed that relaxation time is appreciable, and indicated the desirability of making measurements under steady state conditions. This could be done by rapidly shearing a thin film of viscous oil between unloaded cylindrical surfaces at controlled or measurable temperatures, under conditions permitting the evaluation of the temperature effects. The author is indebted to his colleagues, Mr. R. J. Cornish and Dr. A. F. H. Ward, for helpful discussions on this problem.

References.

- (1) Reynolds, Phil. Trans. p. 162 (1886).
- (2) Edsall, J. T., 'Advances in Colloid Science': Article on "Streaming Birefringence," p. 269. New York (1942).
- (3) Robinson, J. R., Proc. Roy. Soc. A, clxx. p. 531 (1939).
- (4) Farrow, Lowe and Neale, J. Text. Inst. T. xix. p. 18 (1928).
- (5) Bucheim, Stuart and Menz, *Zeit. f. Physik.* cxii. p. 407 (1939).
- (6) Kyropoulos, *Forschung Ingenieurwesen*, ser. 3, p. 287 (1932); J. Chem. Phys. vii. p. 52 (1939).
- (7) Neale, 'Chemistry and Industry,' p. 140 (1937).
- (8) McKee and McKee, Proc. Amer. Soc. Mech. Eng. p. 161 (1929).
- (9) Kingsbury, Mech. Eng. lv. p. 685 (1933).
- (10) Spiers, J. Inst. Auto. Eng. p. vii (Jan. 1941).
- (11) Lamb, 'Hydrodynamics,' 5th edn. p. 555.
- (12) Cornish, Proc. Roy. Soc. A, cxl. p. 227 (1933).
- (13) Fage, Proc. Roy. Soc. A, clxv. p. 501 (1938).
- (14) Hersey, 'Theory of Lubrication.' New York (1936).

LXIX. On the Dimensions of Physical Magnitudes (Third Paper: *Electric and Magnetic Magnitudes*).

By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.*

[Received April 7, 1943.]

I.

IN an earlier paper † the dimensions of electric and magnetic magnitudes were briefly discussed in the light of certain general principles, and a

* Communicated by the Author.

† Phil. Mag. xxxiii. p. 321 (1942).

scheme for such magnitudes was developed far enough to arrive at two dimensional constants which took the place in this scheme of the κ and μ of ordinary theory. Unfortunately, through a careless error in expression, these constants were represented as "equivalent" to κ and μ . This was misleading, flagrantly so with respect to one of the constants which I called μ and stated to be "equivalent to permeability." This constant was introduced in equation (49), viz.,

$$H = \frac{\mu q}{rt},$$

for the magnetic field at the centre of a circular coil carrying a current.

Several correspondents have pointed out the error, but almost, if not quite, all have mistaken its character, and have suggested that μ should be omitted from the equation (or at most made a pure number, without dimensions) because the force on a magnetic pole at the centre of the coil is independent of the medium there. It is true, of course, that the field is independent of the medium, but a dimensional constant is nevertheless necessary in the equation to make it homogeneous, as may be seen by putting in the dimensions of H , q , r and t according to the scheme I adumbrated. What is wrong is the statement that this constant is "equivalent" to permeability. The error is without effect on the validity of the scheme proposed, but it does affect the relation of that scheme to the existing one, and this has proved so serious to the understanding of the principles involved that it seems necessary to develop the scheme more completely, so that its character may be clearly understood. It will then be seen, I think, that it was not built on ignorance of an elementary fact in electro-magnetism, though I make no attempt to excuse an unpardonable piece of carelessness.

II.

According to the principles already set out, the systematic development of any branch of physics, including electro-magnetism, consists in first *defining* the magnitudes to be dealt with, in terms of the processes of measurement by which their values are determined; and then *finding by experiment* the relations between them. In the first part of this programme (forming the definitions) any constants which appear are mere numbers, non-dimensional, which serve merely to control the size of the unit of the magnitude in question; in the second part (finding experimental relations), however, the definitions having been fixed, constants having dimensions will in general occur, and their dimensions will be determined by the equations expressing the experimental results.

For example, if we define force as the rate of change of momentum, we shall have the *defining* dimensional equation

$$[F] = [MLT^{-2}], \quad \dots \dots \dots (1)$$

and if, for convenience, we wish to choose the unit of force so that

$$f = N \frac{d}{dt}(mv), \quad \dots \dots \dots (2)$$

by the quotient of the force and the charge. The dimensions of E are therefore given by

$$[E] = \left[\frac{F}{Q} \right] = [1], \quad (7)$$

i. e., electric field intensity is dimensionless.

(iii) POTENTIAL DIFFERENCE, V .—Let a charge q be moved from a point A to a point B in an electric field. If W is the work done in the process, the potential difference between A and B is defined as $\frac{W}{q}$. The dimensional equation is therefore

$$[V] = \left[\frac{W}{Q} \right] = [L]. \quad (8)$$

(iv) CAPACITY, C .—Measure the charge q on a conductor, and the work required to bring a unit charge from infinity to the conductor (i. e., the potential difference, V , between the conductor and infinity). The capacity of the conductor is then $\frac{q}{V}$. The dimensions of capacity are therefore

$$[C] = \left[\frac{Q}{V} \right] = \left[\frac{F}{L} \right] = [MT^{-2}]. \quad (9)$$

(v) MAGNETIC POLE, p .—Choose any arbitrary standard pole, place the pole to be measured at a fixed position with respect to it, the whole system being in a vacuum, and measure the force between the poles: the magnitude of the unknown pole is then to be taken as proportional to this force. As with electric charge, the dimensions of magnetic pole are

$$[P] = [F] = [MLT^{-2}]. \quad (10)$$

(vi) MAGNETIC FIELD INTENSITY, H .—Place a pole in the field and measure the force acting on it. The strength of the field is then measured by the quotient of the force and the pole. The dimensions of H are then given by

$$[H] = \left[\frac{F}{P} \right] = [1], \quad (11)$$

i. e., magnetic field intensity is dimensionless. In the earlier paper I defined magnetic field independently of pole strength, though in such a way that it was still dimensionless, since I was then concerned merely to arrive by the quickest route at the second dimensional constant taking part in the scheme. In a more complete treatment, however, it seems better to start with pole strength in order to preserve the symmetry between pole and electric charge. The procedures are, however, precisely equivalent so far as the dimensions of the magnitudes are concerned, and if one wishes to avoid the somewhat artificial conception of pole strength it is quite simple to do so.

(vii) ELECTRIC CURRENT, i .—Pass the current through a circular coil containing one turn of wire and having radius r , and measure the magnetic

field H at the centre of the coil. Then the strength of the current is defined by $i=Hr$. The dimensional equation is

$$[I]=[HL]=[L]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

(viii) ELECTROMOTIVE FORCE, E' . Measure the current i in the circuit (or portion of the circuit) and let it be used entirely in producing a quantity of heat H' in time t . Then the E.M.F. is measured by $\frac{JH'}{it}$. The dimensional equation is

$$[E']=\left[\frac{W}{It}\right]=[MLT^{-3}]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

It will be observed that no experimental result is assumed in the foregoing definitions. They are such that the corresponding magnitudes can be seen to be uniquely determined without drawing on any knowledge acquired by experiment. It is assumed, of course, that the quantities can be *recognized*—that a charge, for example, is identifiable by its influence on the standard charge—but no *quantitative* relations are presupposed. For example, the validity of the definition of electric field intensity as force divided by charge may seem to depend on an experimental proof that this quotient is the same whatever charge is used. But actually this is guaranteed by the process of measuring charge, which makes it proportional to the force acting on it in a given field. A similar remark applies to magnetic field intensity and pole strength. The same fact makes the definition of potential difference independent of the assumption that the work done is proportional to the charge used.

An interesting example is the definition of electric current, which may be thought to depend on the experimental fact that the field at the centre of the coil for a given current is inversely proportional to the radius of the coil. This is not so, however. Suppose we have a circuit containing two such coils having different radii, and we send a current through it and measure the magnetic field at the centre of each coil. The definition then does not imply that Hr will be the same for both, and therefore it must not be taken as axiomatic that the current is the same in all parts of a circuit. That, as a matter of fact, is an experimental fact which we shall presently introduce in the form that a current is equivalent to the movement of electric charges round the circuit. It then follows that a steady current must be the same everywhere in the circuit, but that is not required by the definition. Again, if the shape of a circuit is altered so that a coil in it takes a larger radius, the definition does not require that the field at its centre shall be proportionally decreased; that also is a fact of experiment, but it is not required by the definition.

Resistance has been omitted from the list of definitions because, although it can be defined in the ordinary way, the definition is useless without the experimental fact of Ohm's law that in certain circumstances

its value is independent of the current and E.M.F. used. We shall therefore introduce it as an experimentally determined constant.

The foregoing list of quantities, though not complete, is sufficient for our purpose. We now proceed to the experimentally established relations between them, again limiting ourselves to a number sufficient to make the general principles clear. The dimensional constant which, as already explained, occurs in each will be represented by a Greek letter.

(i) *A current of electricity in a circuit is equivalent to the motion of electric charges round the circuit, and its measure is proportional to the charge which passes any point of the circuit in a given time.* We have, therefore,

$$i = \alpha \frac{q}{t}, \quad (14)$$

or, dimensionally,

$$[\alpha] = [L T F^{-1}] = [M^{-1} T^3]. \quad (15)$$

It follows from this result, together with the definitions, that E.M.F. is equivalent to potential difference, and we can easily verify that

$$[V] = [\alpha E']. \quad (16)$$

It follows also, as already explained, that a steady current has the same value at all parts of its circuit.

(ii) *The inverse square law of electric charges, viz.,*

$$f = \frac{q_1 q_2}{\beta r^2} (17)$$

$$\left. \begin{array}{l} \text{Dimensionally,} \quad [F] = [F^2 \beta^{-1} L^{-2}], \\ \text{or} \quad \quad \quad [\beta] = [F L^{-2}] = [M L^{-1} T^{-2}]. \end{array} \right\} (18)$$

(iii) *The inverse square law of magnetic poles, viz.,*

$$f = \frac{p_1 p_2}{\epsilon r^2}, \quad (19)$$

whence, as with electric charges,

$$[\epsilon] = [M L^{-1} T^{-2}] = [\beta]. \quad (20)$$

(iv) *A current i , placed in a magnetic field H , experiences a force given by*

$$f = \zeta l H, \quad (21)$$

where l is the length of the wire carrying the current. Hence

$$[F] = [\zeta L^2], \quad (22)$$

$$\text{so that} \quad [\zeta] = [F L^{-2}] = [M L^{-1} T^{-2}] = [\beta]. \quad (23)$$

(v) *Ohm's law, viz.,*

$$E' = \rho i, \quad (24)$$

$$\text{whence} \quad [\rho] = [E' I^{-1}] = [M T^{-2}] = [\alpha^{-1}]. \quad (25)$$

Additional experimental results yield nothing further. We have two dimensional constants, α and β , having dimensions $[M^{-1}T^3]$ and $[ML^{-1}T^{-2}]$ respectively, and these are the representatives on this scheme of the μ and κ of ordinary theory. Corresponding to the relation

$$[\mu^{-\frac{1}{2}} \kappa^{-\frac{1}{2}}] = [LT^{-1}], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

we have now

$$\alpha^{-1}\beta^{-1}=[LT^{-1}]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

α and β were called μ and κ in my earlier paper, and so were falsely made to appear identical with permeability and dielectric constant. It may now be seen more clearly in what sense they correspond to permeability and dielectric constant, and how inappropriate, without being entirely false, it was to say that they were "equivalent" thereto. In such a context the word "equivalent" could scarcely fail to suggest that I was falsely asserting that the magnetic field depended on the *material* medium in the region in question, whereas actually my μ was equivalent to permeability only in a much more general sense.

Just as no further experimental results increase the number of independent dimensional constants, so no incorporation of such results in the definitions decreases the number. If, for example, we make use of the first result to *define* current as q/t , thus making α unnecessary, our previous definition (vii) then becomes an experimental result which re-introduces it. This, in fact, was the way in which α was introduced in the earlier paper. The number of dimensional constants involved in the scheme is fixed at 2, and in this respect it resembles the ordinary scheme. It is thus only in a very general sense that α and β can be said to be equivalent to μ and κ .

IV.

It is important to understand the relation of the foregoing scheme to that in current use. In the first place, the two schemes are precisely equivalent so far as experiment is concerned. If the sizes of the units are chosen (as they obviously can be) to be equal to those now adopted, every equation will be numerically the same in the two schemes. There is, therefore, no ground whatever for distinguishing between them with respect to their ability to represent all known facts.

Secondly, whereas the usual scheme involves two distinct systems of measurement—the electrostatic and electromagnetic systems—the scheme outlined above involves only one. This is not, of course, merely a matter of *size* of unit—the unit of any quantity can be of any size in either of the schemes—but it concerns the *dimensions*, i. e., the “physical nature” of the quantities according to traditional theory, or the “process of measurement” according to the view adopted here. There is thus no need to “identify” different dimensions of the “same” quantity, as is usually done.

Thirdly, whereas in the current scheme the dimensions of μ and κ are "unknown" (*i. e.* we can make them what we wish provided that the

product has the dimensions $L^{-2}T^3$), the dimensions of α and β are determined by the manner in which they appear in the equations.

Fourthly, in the present scheme the definition of each measurement is separately and unambiguously stated, whereas in the other the definitions of electric charge and magnetic pole are inextricably confused with those of κ and μ , respectively, and, moreover, measurements (such as that of the electromagnetic unit of quantity of electricity, for instance) are presumed which are never made.

In each of these differences the scheme proposed above clearly has the advantage, and since it is equivalent to the other in its power of representing all our experimental knowledge, it should, I think, be given the preference. Dr. Norman Campbell has recently * given five "ends" to which the assignment of dimensions is a "means." There is a sixth, which is to my mind the most important of all—namely, that the clear and unambiguous assignment of dimensions is the most direct way of ensuring that one understands precisely what the subject involves, what it postulates, what it assumes, and what experimental results it includes (see, for example, next section). If one is forced to say that the dimensions of a quantity are "unknown," that is tantamount to saying that one does not understand what that quantity means or what its symbol stands for. That can never be justified. If we do not understand a symbol we have no right to use it.

On the other hand, it would be contrary to the whole meaning of dimensions as I understand them to say that the above scheme is the "correct" one and that no other is to be permitted. There are a number of possible schemes, each as valid as any other, in no two of which the dimensions assigned to quantities bearing the same name are the same. For example, it would be possible to define "current" in terms of the force between two wires carrying currents. A current would be measured by bringing it into a fixed position with respect to a standard current, and measuring the force of attraction or repulsion. The dimensions of current would then be those of force, viz. $[MLT^{-2}]$, and the *magnetic* effects of currents would not appear explicitly in the definitions but would be discovered experimentally. Such a scheme would be perfectly legitimate because it would be capable of describing all known phenomena unambiguously, but the dimensions of the quantities measured and of the dimensional constants would be very different from those given above.

It might be added that in a strictly logical development of the subject, in which a quantity is known only by the effect which gives its measure, alternative dimensions for the same quantity would not be possible. We should not be able to define current alternatively by its magnetic effect or by its effect on another current; if "current" is measured by a magnetic effect, that which is measured by its self-attraction should be given another name. We should then say that alternative "concepts"

* Phil. Mag. xxxiii. p. 398 (1942).

are possible in terms of which the content of electro-magnetism could be described, each concept having its own unique dimensions. To adopt this course, however, would take us further from current practice than is necessary or desirable at present, and we shall therefore not do so.

The experimental facts of a science, as expressed by the dimensional equations, form a framework which it is customary to clothe with a theory. The traditional formulæ have been draped with the ether theory, according to which μ and κ are some unspecified properties of the ether which are modified by the presence of matter and which are related through their product to the velocity with which the ether transmits transverse waves. Such a theory is not readily adaptable to the scheme proposed above, for although β might be naturally regarded as a property of a medium, α (the current equivalent to the passage of unit quantity of electricity in unit time) is less readily so interpreted. The ether theory, however, has decidedly lost caste since the time of Maxwell, and the more direct expression of the facts provided by the above scheme is an invitation to anyone so minded to clothe it with a more satisfactory theory. None is attempted here.

V.

It will be helpful to give one example of the way in which dimensional formulæ as here interpreted make clear precisely what is involved in a physical equation; the example which follows relates particularly to the significance of dimensional constants as an indication that an *experimental* result is involved.

The "fine-structure constant," $\frac{ch}{2\pi e^2}$, is frequently referred to as "dimensionless"; for example, Eddington* states that when we say that its value is 137 we are "eliminating all reference to our artificial standards." It is clear, however, that this is not so. According to either the customary scheme or that outlined here, this constant has dimensions— κ^{-1} in the former case and β^{-1} in the latter. It is, indeed, directly obvious that we may change the unit of e (*e. g.*, to that charge which repels an equal and opposite charge at unit distance with a force of 2 units) without at all changing the units involved in c and h , and the value of the fine-structure constant, therefore, does depend on our artificial standards.

The matter, of course, is easily adjusted by defining the constant as $\frac{ch\kappa}{2\pi e^2}$ or $\frac{ch\beta}{2\pi e^2}$, but this is, in fact, a fundamental change, for the appearance of κ or β is an infallible indication that the results of *experiment* have been drawn upon; it is inherently impossible, the universally accepted meanings of the quantities involved being granted, for the dimensionless

* 'New Pathways in Science,' p. 232.

fine-structure constant to be derived from pure reason. (This has nothing to do, of course, with the fact that the value of κ or β varies with the medium. If, like the gravitational constant, γ —compare equation (3)—it were independent of the medium, its necessary appearance would still indicate dependence on experiment.)

It follows, then, that in any procedure, such as that followed by Eddington, by which the fine-structure constant is derived *a priori*, the symbols must have different meanings from those generally understood. We are therefore hopelessly at cross-purposes with him if we challenge his claim in the belief that he is referring to the quantities with which we are familiar. There are, in fact, two ways in which the fine-structure constant can conceivably be made into a dimensionless relation between physical magnitudes. Let us consider them in turn.

The first would be realized if κ or β were interpreted not as a "dimensional constant" but as something derived directly from an experiment independent of those by which c , h and e are determined. As an example, suppose a pioneer experimenter (*e. g.* Galileo) discovered that for small swings the time of vibration of a simple pendulum was proportional to the square root of its length. He would express this result by the equation

$$t^2 = \frac{l}{\psi}, \quad (28)$$

where ψ would be a dimensional constant. He might next find that a freely falling body at the same place had a constant acceleration proportional to the ψ already found. He would then be able to *define* ψ , independently of (28), as this acceleration, and call it g . His equation for the pendulum would then be

$$t^2 = Nlg^{-1}, \quad (29)$$

where N is the constant of proportionality, and it is clear, on inserting the dimensions, that N must be dimensionless. He would thus have arrived at a dimensionless relation, $\frac{t^2 g}{l}$, between three *independently* measured quantities— l , the length of the pendulum; t , its time of swing; and g , the acceleration of a neighbouring freely falling body. The way would then be open for him to search for a point of view from which the numerical value of his constant could be derived by pure reason.

We know of no method of measuring κ or β directly in this way. It is at present a constant analogous to ψ , not to g , and neither Eddington nor anyone else has indicated how it can be measured independently. (We could, it is true, draw on another of its occurrences in electromagnetism, such as that indicated by equations (21) to (23), but this would introduce currents and magnetic fields into the relation, which would alter the nature of the fine-structure constant too radically for our present considerations.) We turn, therefore, to the second way in

which the fine-structure constant may be regarded as capable of *a priori* determination.

This is illustrated by the identification of inertial and gravitational mass in general relativity. With Newtonian definitions the appearance of the dimensional gravitational constant, γ , assures us that equation (3) cannot be derived without experiment, and, as already explained, if we define mass gravitationally, then the inertial property of mass embodied in (2) becomes an experimental result. Hence we cannot release ourselves from experiment; we can only choose what experiment we will make. Einstein escaped from this dilemma by abandoning the Newtonian point of view and its definitions entirely. It is not strictly true to say that he identified inertial and gravitational mass. He ignored them, and introduced a single new quantity which served the purposes of both. It is, in fact, literally nonsense to speak, as relativists do, of a mass of so many metres, but the nonsense is not in the theory of relativity but in the use of the word "mass" for a quantity which is a function of measurements with metre rods. (This is, incidentally, very similar to my nonsense in speaking of my μ as equivalent to permeability.)

It is along these lines that Eddington seeks to derive the value of the fine-structure constant by pure reason. This is no place to discuss the validity of his deduction, or the degree to which it is independent of experience of some kind, but what should be made clear is that, valid or not and independent of experience or not, what he arrives at is not the quantity $\frac{ch\kappa}{2\pi e^2}$ but something else which he claims fulfils the same function. I do not think that this can be shown in any simple way without the use of a consistent and non-metaphysical theory of dimensions.

VI.

In my second paper on this subject*, in which I considered some criticisms by Professor Wilson of the first paper, I stated that I found his comments on electrical and magnetic magnitudes "completely unintelligible." I must slightly modify this now, since the realization of my own *faux pas* suggests that it has led him to attribute to me the elementary error referred to in the first section of this paper. If that is so I can understand a portion of his comments—namely, the exclamation mark in line 12 on p. 635. As soon as I attempt to advance, however, the veil descends again and the argument becomes impenetrable. How he can assert his equation (15), adopting, as he does, my dimensions for the magnitudes involved, or write $[M]=[T^3]$ after M, L and T have been chosen as fundamental magnitudes, and proceed without realizing that he has reached an absurdity, I cannot imagine. I should be very pleased if I could reach some understanding with him on this subject, but I can find no basis whatever for it in anything that he has yet written.

* Phil. Mag. xxxiii. p. 692 (1942). Wilson's paper appeared in Phil. Mag. xxxiii. p. 631 (1942).

LXX. *An Integral related to the Radiation Integrals.*

By E. O. POWELL*.

[Received February 1, 1943.]

THE integrals $\text{Ri}_n(x) = \int_x^\infty \frac{d\xi}{\xi^n(e^\xi - 1)}$

and $\int_0^x \frac{\xi^n d\xi}{e^\xi - 1} = \Gamma(1+n)\zeta(1+n) - \text{Ri}_{-n}(x)$
 $(n > 0)$

are of frequent occurrence in astrophysics and quantum mechanics. Airey (1938) tabulates $\text{Ri}_n(x)$ for $n=0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 2$; and Debye (1913)

gives a short table of $\frac{1}{x} \int_0^x \frac{\xi d\xi}{e^\xi - 1} = \frac{\pi^2}{6x} - \frac{1}{x} \text{Ri}_{-1}(x)$.

The integral $\int_1^x \frac{\log \xi d\xi}{\xi - 1} = \text{Ri}_{-1}\left(\log \frac{1}{x}\right) - \frac{\pi^2}{6}$,

was recently encountered in a physical problem (Finn and Powell, 1942), and in view of the possibly wider interest given to it by this relation, a table of values was computed. This integral will be denoted here by $\text{Rl}(x)$.

The values were obtained from three series and a number of functional equations satisfied by $\text{Rl}(x)$:—

(i) Writing $\log \xi$ as $\log \{1 + (\xi - 1)\}$ and expanding :

$$\text{Rl}(x) = \sum_1^\infty (-1)^{n+1} (x-1)^n \cdot n^{-2} \quad \dots \dots \dots \quad (I)$$

$(2 \geq x \geq 0),$

$$(ii) \text{Rl}(x) = \log x \log (x-1) - \int_1^x \frac{\log \xi d\xi}{\xi} - \int_1^x \frac{\log (1-1/\xi) d\xi}{\xi}$$

$$= \log x \log (x-1) - \frac{1}{2}(\log x)^2 + \frac{\pi^2}{6} - \sum_1^\infty x^{-n} n^{-2} \quad \dots \dots \dots \quad (II)$$

$(x \geq 1),$

$$(iii) \text{Rl}(x) = \int_1^{\frac{1}{x}} \frac{\log \xi d\xi}{\xi(1-\xi)}, \text{ substituting } \frac{1}{\xi} \text{ for } \xi,$$

$$= \int_1^{\frac{1}{x}} \left\{ \frac{\log \xi}{\xi} + \frac{\log \xi}{1-\xi} \right\} d\xi$$

$$= \frac{1}{2}(\log x)^2 - \log \frac{1}{x} \log \left(\frac{1}{x} - 1 \right) + \frac{1}{2}(\log x)^2 - \frac{\pi^2}{6} + \sum_1^\infty x^n n^{-2}.$$

(from II.)

* Communicated by the Author.

Hence
$$\text{Rl}(x) = \log x \log (1-x) - \frac{\pi^2}{6} + \sum_1^{\infty} x^n n^{-2} \dots \dots \dots \text{(III)}$$

 $(0 \leq x \leq 1).$

(iv) Two equations enable all the function values to be computed from those in the ranges $0 < x < 1$ and $1 < x < 2$:

From (iii) above
$$\text{Rl}(x) = \int_1^{\frac{1}{x}} \left\{ \frac{\log \xi}{\xi} + \frac{\log \xi}{1-\xi} \right\} d\xi.$$

Hence
$$\text{Rl}(x) + \text{Rl}\left(\frac{1}{x}\right) = \frac{1}{2}(\log x)^2. \dots \dots \dots \text{(IV)}$$

Substituting ξ^2 for ξ in $\text{Rl}(x^2)$ and separating into partial fractions

$$\begin{aligned} \frac{1}{2}\text{Rl}(x^2) &= \text{Rl}(x) + \int_1^x \frac{\log \xi d\xi}{\xi+1} \\ &= \text{Rl}(x) + \log x \log (x+1) - \int_2^{x+1} \frac{\log \xi d\xi}{\xi-1}. \end{aligned}$$

From (I), $\text{Rl}(2) = \frac{\pi^2}{12}$, hence

$$\text{Rl}(x+1) - \text{Rl}(x) = \log x \log (x+1) + \frac{\pi^2}{12} - \frac{1}{2}\text{Rl}(x^2). \dots \dots \text{(V)}$$

From (I) and (II)
$$\text{Rl}\left(1 + \frac{1}{x}\right) = \sum_1^{\infty} (-1)^{n+1} x^{-n} n^{-2}$$

$$\sum_1^{\infty} x^{-n} n^{-2} = \log x \log (x-1) - \frac{1}{2}(\log x)^2 + \frac{\pi^2}{6} - \text{Rl}(x);$$

subtracting, and replacing $\sum_1^{\infty} x^{-2n} n^{-2}$ by its value as given by (II), then eliminating $\text{Rl}(x)$ with (V) :

$$\text{Rl}(x+1) + \text{Rl}\left(\frac{1}{x} + 1\right) = \frac{1}{2}(\log x)^2 + \frac{\pi^2}{6}. \dots \dots \text{(VI)}$$

(v) In the range $1 \geq x \geq 0$, from (I) and (III)

$$\text{Rl}(1-x) = - \sum_1^{\infty} x^n n^{-2} = -\text{Rl}(x) + \log x \log (1-x) - \frac{\pi^2}{6},$$

$$\text{Rl}(x) + \text{Rl}(1-x) = \log x \log (1-x) - \frac{\pi^2}{6}, \dots \dots \dots \text{(VII)}$$

and from (I)

$$\begin{aligned} \text{Rl}(1-x) + \text{Rl}(1+x) &= - \sum_1^{\infty} x^n n^{-2} + \sum_1^{\infty} (-1)^{n+1} x^n n^{-2} \\ &= - \frac{1}{2} \sum_1^{\infty} x^{2n} n^{-2} \end{aligned}$$

$$\text{Rl}(1-x) + \text{Rl}(1+x) = \frac{1}{2}\text{Rl}(1-x^2). \dots \dots \dots \text{(VIII)}$$

It is hoped that the tabular values will be found correct to six decimal places; the seventh should not usually be in error by more than one unit. It should be noted that $\text{Rl}(x)$ is negative for $0 < x < 1$, and that

$$\text{Rl}(x) = \int_1^x \frac{\log \xi d\xi}{\xi - 1}.$$

x .	$\text{Rl}(x)$.	δ^2 .	x .	$\text{Rl}(x)$.	δ^2 .
0.00	-1.6449 341		0.40	-0.7275 863	-1 621
1	1.5886 274	-134 789	1	7123 950	1 573
2	1.5457 997	49 270	2	6973 611	1 526
3	1.5078 990	31 275	3	6824 797	1 482
4	1.4731 259	22 811	4	6677 466	1 441
0.05	-1.4406 338	-17 866	0.45	-0.6531 576	-1 401
6	1.4099 283	14 622	6	6387 087	1 363
7	1.3806 850	12 334	7	6243 961	1 327
8	1.3526 752	10 635	8	6102 161	1 292
9	1.3257 287	9 324	9	5961 654	1 259
0.10	-1.2997 147	-8 284	0.50	-0.5822 405	-1 228
1	1.2745 292	7 440	1	5684 384	1 197
2	1.2500 876	6 741	2	5547 561	1 168
3	1.2263 201	6 153	3	5411 906	1 140
4	1.2031 680	5 653	4	5277 391	1 114
0.15	-1.1805 811	-5 223	0.55	-0.5143 990	-1 088
6	1.1585 165	4 847	6	5011 677	1 064
7	1.1369 366	4 518	7	4880 428	1 040
8	1.1158 084	4 228	8	4750 219	1 017
9	1.0951 031	3 968	9	4621 027	995
0.20	-1.0747 946	-3 738	0.60	-0.4492 830	- 974
1	1.0548 598	3 528	1	4365 607	954
2	1.0352 780	3 340	2	4239 338	934
3	1.0160 301	3 169	3	4114 003	915
4	0.9970 991	3 012	4	3989 583	897
0.25	-0.9784 694	-2 871	0.65	-0.3866 059	- 879
6	9601 267	2 738	6	3743 415	862
7	9420 578	2 618	7	3621 633	846
8	9242 506	2 506	8	3500 696	829
9	9066 941	2 402	9	3380 589	814
0.30	-0.8893 776	-2 305	0.70	-0.3261 295	- 799
1	8722 917	2 216	1	3142 800	784
2	8554 274	2 132	2	3025 090	770
3	8387 763	2 054	3	2908 150	757
4	8223 305	1 980	4	2791 967	743
0.35	-0.8060 827	-1 911	0.75	-0.2676 526	- 730
6	7900 260	1 846	6	2561 817	718
7	7741 540	1 785	7	2447 825	706
8	7584 604	1 727	8	2334 539	694
9	7429 396	1 674	9	2221 947	683
0.40	-0.7275 863	-1 621	0.80	-0.2110 038	- 671

x .	$Rl(x)$.	δ^2 .	x .	$Rl(x)$.	δ^2 .
0.80	-0.2110 038	- 671	1.20	0.1908 001	- 391
1	1998 800	661	1	1998 967	387
2	1888 223	650	2	2089 546	383
3	1778 295	640	3	2179 742	378
4	1669 008	630	4	2269 560	375
0.85	-0.1560 350	- 620	1.25	0.2359 003	- 370
6	1452 313	611	6	2448 076	366
7	1344 886	602	7	2536 783	363
8	1238 060	592	8	2625 127	358
9	1131 827	584	9	2713 113	356
0.90	-0.1026 178	- 575	1.30	0.2800 743	- 350
1	0921 104	567	1	2888 023	347
2	816 596	559	2	2974 956	345
3	712 647	551	3	3061 544	340
4	609 248	543	4	3147 792	336
0.95	-0.0506 393	- 535	1.35	0.3233 704	- 334
6	404 073	528	6	3319 282	330
7	302 281	521	7	3404 530	327
8	201 009	514	8	3489 451	323
9	100 251	507	9	3574 049	321
1.00	+0.0000 000	- 500	1.40	0.3658 326	- 317
1	99 751	493	1	3742 286	315
2	199 009	487	2	3825 931	310
3	297 780	481	3	3909 266	308
4	396 070	475	4	3992 293	306
1.05	+0.0493 885	- 468	1.45	0.4075 014	- 302
6	591 232	462	6	4157 433	299
7	688 117	457	7	4239 553	298
8	784 545	452	8	4321 375	293
9	880 521	445	9	4402 904	291
1.10	0.0976 052	- 440	1.50	0.4484 142	- 289
1	1071 143	434	1	4565 091	285
2	1165 800	431	2	4645 755	284
3	1260 026	424	3	4726 135	281
4	1353 828	419	4	4806 234	278
1.15	0.1447 211	- 415	1.55	0.4886 055	- 276
6	1540 179	409	6	4965 600	274
7	1632 738	406	7	5044 871	270
8	1724 891	400	8	5123 872	268
9	1816 644	396	9	5202 605	266
1.20	0.1908 001	- 391	1.60	0.5281 072	- 264

x .	$Rl(x)$.	δ^2 .	x .	$Rl(x)$.	δ^2 .
1.60	0.5281 072	— 264	2.00	0.8224 670	— 772
1	5359 275	263	2	8362 915	762
2	5437 215	259	4	8500 398	751
3	5514 896	257	6	8637 130	741
4	5592 320	255	8	8773 121	731
1.65	0.5669 489	— 253	2.10	0.8908 381	— 721
6	5746 405	250	2	9042 920	712
7	5823 071	249	4	9176 747	702
8	5899 488	247	6	9309 872	693
9	5975 658	244	8	9442 304	683
1.70	0.6051 584	— 243	2.20	0.9574 053	— 676
1	6127 267	242	2	9705 126	666
2	6202 709	238	4	9835 533	658
3	6277 913	237	6	9965 282	649
4	6352 880	235	8	1.0094 382	642
1.75	0.6427 612	— 232	2.30	1.0222 840	— 634
6	6502 112	232	2	1.0350 664	625
7	6576 380	229	4	1.0477 863	619
8	6650 419	228	6	1.0604 443	610
9	6724 230	225	8	1.0730 413	604
1.80	0.6797 816	— 225	2.40	1.0855 779	— 596
1	6871 177	222	2	1.0980 549	589
2	6944 316	220	4	1.1104 730	582
3	7017 235	219	6	1.1228 329	576
4	7089 935	217	8	1.1351 352	569
1.85	0.7162 418	— 215	2.50	1.1473 806	— 562
6	7234 686	215	2	1.1595 698	556
7	7306 739	212	4	1.1717 034	549
8	7378 580	211	6	1.1837 821	544
9	7450 210	209	8	1.1958 064	538
1.90	0.7521 631	— 207	2.60	1.2077 769	— 531
1	7592 845	207	2	1.2196 943	525
2	7663 852	204	4	1.2315 592	520
3	7734 655	203	6	1.2433 721	514
4	7805 255	202	8	1.2551 336	510
1.95	0.7875 653	— 200	2.70	1.2668 441	— 503
6	7945 851	198	2	1.2785 043	498
7	8015 851	198	4	1.2901 147	492
8	8085 653	196	6	1.3016 759	488
9	8155 259	195	8	1.3131 883	482
2.00	0.8224 670	— 193	2.80	1.3246 525	— 477

x .	$\text{Rl}(x)$.	δ^2 .	x .	$\text{Rl}(x)$.	δ^2 .
2.80	1.3246 525	— 477	3.60	1.7485 815	— 331
2	1.3360 690	474	2	1.7584 183	327
4	1.3474 381	467	4	1.7682 224	326
6	1.3587 605	463	6	1.7779 939	322
8	1.3700 366	459	8	1.7877 332	320
2.90	1.3812 668	— 453	3.70	1.7974 405	— 318
2	1.3924 517	450	2	1.8071 160	315
4	1.4035 916	444	4	1.8167 600	312
6	1.4146 871	441	6	1.8263 728	310
8	1.4257 385	436	8	1.8359 546	308
3.00	1.4367 463	— 432	3.80	1.8455 056	— 305
2	1.4477 109	428	2	1.8550 261	303
4	1.4586 327	423	4	1.8645 163	301
6	1.4695 122	420	6	1.8739 764	297
8	1.4803 497	416	8	1.8834 068	296
3.10	1.4911 456	— 411	3.90	1.8928 076	— 294
2	1.5019 004	409	2	1.9021 790	291
4	1.5126 143	404	4	1.9115 213	289
6	1.5232 878	400	6	1.9208 347	288
8	1.5339 213	397	8	1.9301 193	285
3.20	1.5445 151	— 393	4.00	1.9393 754	— 282
2	1.5550 696	390	2	1.9486 033	281
4	1.5655 851	385	4	1.9578 031	279
6	1.5760 621	383	6	1.9669 750	277
8	1.5865 008	379	8	1.9761 192	274
3.30	1.5969 016	— 376	4.10	1.9852 360	— 273
2	1.6072 648	373	2	1.9943 255	270
4	1.6175 907	369	4	2.0033 880	268
6	1.6278 797	366	6	2.0124 237	268
8	1.6381 321	362	8	2.0214 326	264
3.40	1.6483 483	— 360	4.20	2.0304 151	— 263
2	1.6585 285	356	2	2.0393 713	262
4	1.6686 731	354	4	2.0483 013	259
6	1.6787 823	351	6	2.0572 054	257
8	1.6888 564	347	8	2.0660 838	256
3.50	1.6988 958	— 345	4.30	2.0749 366	— 253
2	1.7089 007	342	2	2.0837 641	253
4	1.7188 714	338	4	2.0925 663	251
6	1.7288 083	336	6	2.1013 434	248
8	1.7387 116	334	8	2.1100 957	247
3.60	1.7485 815	— 331	4.40	2.1188 233	— 245

x .	$Rl(x)$.	δ^2 .	x .	$Rl(x)$.	δ^2 .
4.40	2.1188 233	— 245	5.20	2.4494 198	— 190
2	2.1275 264	244	2	2.4572 611	190
4	2.1362 051	241	4	2.4650 834	189
6	2.1448 597	241	6	2.4728 868	187
8	2.1534 902	239	8	2.4806 715	187
4.50	2.1620 968	— 237	5.30	2.4884 375	— 185
2	2.1706 797	235	2	2.4961 850	184
4	2.1792 391	235	4	2.5039 141	183
6	2.1877 750	232	6	2.5116 249	182
8	2.1962 877	231	8	2.5193 175	181
4.60	2.2047 773	— 230	5.40	2.5269 920	— 180
2	2.2132 439	228	2	2.5346 485	179
4	2.2216 877	226	4	2.5422 871	179
6	2.2301 089	225	6	2.5499 078	177
8	2.2385 076	224	8	2.5575 108	176
4.70	2.2468 839	— 222	5.50	2.5650 962	— 175
2	2.2552 380	220	2	2.5726 641	174
4	2.2635 701	220	4	2.5802 146	173
6	2.2718 802	218	6	2.5877 478	172
8	2.2801 685	216	8	2.5952 638	172
4.80	2.2884 352	— 216	5.60	2.6027 626	— 170
2	2.2966 803	214	2	2.6102 444	170
4	2.3049 040	212	4	2.6177 092	168
6	2.3131 065	211	6	2.6251 572	168
8	2.3212 879	210	8	2.6325 884	166
4.90	2.3294 483	— 209	5.70	2.6400 030	— 166
2	2.3375 878	207	2	2.6474 010	165
4	2.3457 066	206	4	2.6547 825	165
6	2.3538 048	205	6	2.6621 475	163
8	2.3618 825	204	8	2.6694 962	162
5.00	2.3699 398	— 202	5.80	2.6768 287	— 162
2	2.3779 769	201	2	2.6841 450	160
4	2.3859 939	200	4	2.6914 453	160
6	2.3939 909	199	6	2.6987 296	159
8	2.4019 680	198	8	2.7059 980	158
5.10	2.4099 253	— 196	5.90	2.7132 506	— 157
2	2.4178 630	195	2	2.7204 875	157
4	2.4257 812	194	4	2.7277 087	156
6	2.4336 800	193	6	2.7349 143	155
8	2.4415 595	192	8	2.7421 044	154
5.20	2.4494 198	— 190	6.00	2.7492 791	— 154

the second (central) differences are negative throughout. These differences are sufficient to give seven-figure accuracy when x is greater than 0.2, six-figure between 0.1 and 0.2; below this limit they increase very rapidly.

I wish to express my thanks to Mr. O. G. Sutton and Dr. W. G. Bickley for their interest and advice in the presentation of this table.

References.

- AIREY, J. R. (1938). The Radiation Integrals. *Phil. Mag.* (7) xxv. pp. 273-282.
 DEBYE, P. (1914). Interferenz von Röntgenstrahlen und Wärmebewegung. *Ann. Phys., Lpz.* xliii. pp. 49-54.
 FINN, S. R. and POWELL, E. O. (1942). The Chemical and Physical Investigation of Germicidal Aerosols. II: The Aerosol Centrifuge. *J. Hyg., Camb.* xlii. pp. 354-364.

The Portslade Research Laboratories,
 Portslade, Sussex.

LXXI. The Duple.

By A. F. DUFTON, M.A., D.I.C.*

[Received May 14, 1943.]

IN a paper describing proposals for a new lighting code⁽¹⁾, Mr. H. C. Weston suggests that recommended values of illumination should, in future, form a true geometrical series, except for deviations necessary to avoid fractions. Mr. Weston envisages a range of thirteen effectively different illuminations, forming a series whose common ratio is 1.414; and, in order to classify tasks on a broader basis, he divides the range into six sub-ranges, whose limits have a common ratio of 2, and remarks that the category limits then form the geometrical series, 2, 2², 2³, . . . 2⁷ ft.c.

It seems pertinent to point out that the question of logarithmic units was raised some ten years ago by Dr. A. H. Davis⁽²⁾, and that the advantages of the simple scale of two have recently been adumbrated by Mr. Peter Simple in a note "The Duple: a logarithmic unit⁽³⁾." Dr. Davis proposed that a 10-fold change in any quantity should be called a change of one *brig*, a unit which he subdivided to obtain a more convenient *decibrig*. That these units have not become current is ascribed by Mr. Simple not only to the fact that the brig itself is rather large, but also to the somewhat involved conception inherent in the decibrig.

* Communicated by the Author.

Remarking that an appropriate name for a binary change would help, he proposes that the necessary connotation should be given to the word *duple* and that the portmanteau-word *tootha* * should be legitimized. Eight-fold is 3-duple; and a 16-fold change or difference, a change of 4 duples. In terms of acoustic intensity, one duple equals 3.01 decibels: an increase of 15 phons is a 5-duple change in loudness.

As illumination levels are energy levels, Mr. Weston suggests that the difference between any two levels can, if desired, be expressed in decibels, and that code illuminations can be expressed in dbs. above a reference level of 1 ft.c. Mr. Simple's proposal appears to be less complicated and no reference level need be cited: 32 foot-candles is tootha 5 ft.c.; but it is 15.05 dbs. above 1 ft.c. Tootha 5 ft.c. is usually written 2^5 ft.c.; the form $2\backslash 5$ ft.c., suggested by Mr. Simple, has typographic merit and is advantageous when the exponent is not an integer and when values are tabled.

References.

- (1) Trans. Ill. Eng. Soc. viii. p. 17 (1943).
- (2) Physical Society, 'Report on a Discussion on Audition,' p. 136 (1931).
- (3) 'Science,' xcv. p. 45 (1942).

The King's Lodge,
Hunton Bridge, Hertfordshire.

LXXII. Notes on Refraction at a Plane Interface and by a (Triangular) Prism.

By ERIC J. IRONS, Ph.D., Queen Mary College, c/o King's College,
Cambridge †.

[Received May 5, 1943.]

I. Refraction and Deviation of Rays at a Plane Interface.

A well-known method for obtaining simultaneous values of i_1 and i_2 (the angles of incidence and of refraction at a plane interface) is to draw two concentric circles of radii n_1 and n_2 (the refractive indices of the materials separated by the interface) and then, by erecting a normal to the interface to pass through E (the point of intersection of the production of the incident ray with the circle of radius n_1) and cut the circle of radius n_2 at G, the direction of the refracted ray is found by joining the point of incidence (O) to G (see fig. 1). This construction is readily justified by noting from the triangle OEG that $n_1/\sin i_2 = n_2/\sin i_1$.

* Tootha=two to the power of.

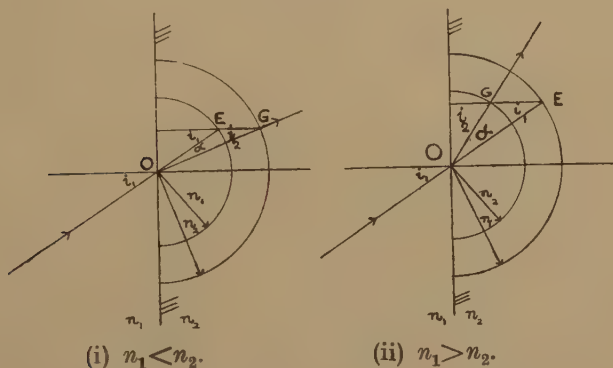
† Communicated by the Author.

What does not seem to be so well realized is that the construction may be utilized :—

(1) To find n_2 given n_1 and some experimentally determined simultaneous values of i_1 and i_2 .

Let us suppose that the refracted paths of rays incident at O at various angles have been traced by some means, then, through the point E, where any one incident ray cuts the circle of radius n_1 , draw EG, the normal to the interface, and produce it to meet OG, the corresponding refracted ray, in G. If this is repeated for the other rays, the value of n_2 is determined by the radius of the arc of centre O about which all such points as G are most evenly distributed.

Fig. 1.



(2) To investigate the deviation produced by refraction.

The deviation produced when the ray whose direction is OE is refracted is shown in the figure by the angle α , which is related to i_1 by

$$\sin \alpha = (EG) \cdot (\sin i_1) / n_2.$$

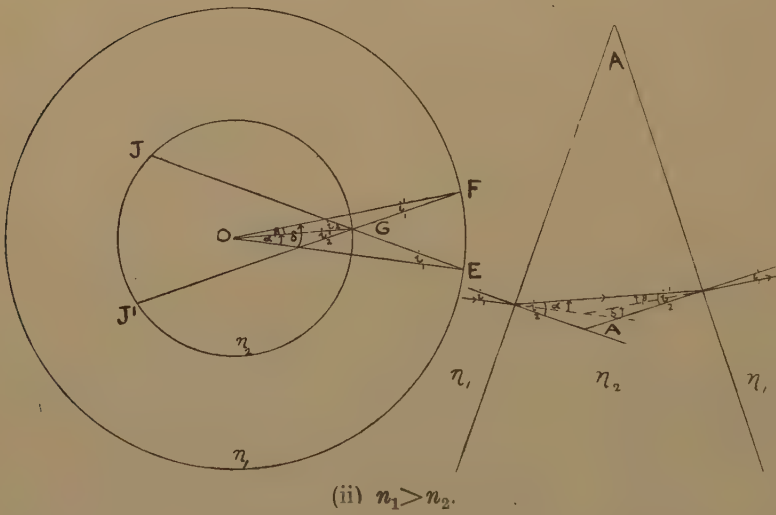
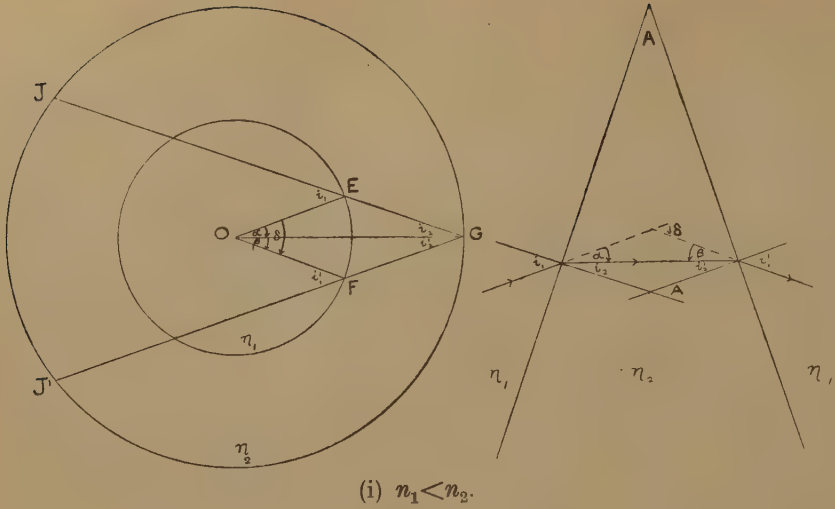
As both $\sin i_1$ and EG increase as i_1 increases, it follows that α increases with i_1 : it may be shown similarly to increase when i_2 increases.

II. The Minimum Deviation Property of a Prism.

This property may be demonstrated very readily by applying the result of the last section*. In figs. 2 and 3 are shown the principal planes of the prisms and the required result is obtained by the consideration of the deviations (δ) produced by the two refractions.

* Note added in proof.—This I find is common practice, see, e. g., Ramsey's 'Elementary Geometrical Optics' (G. Bell and Sons, Ltd.). § 42: in § 43 of the same book, and also for $n_1 < n_2$, is a diagram similar to fig. 2 (i).

Fig. 2.



$$A > i'_2, \text{ i. e., } A = i_2 + i'_2.$$

$$\begin{aligned} \delta &= \alpha + \beta = (i_1 - i_2) + (i'_1 - i'_2) = (i_1 + i'_1) - A & \delta &= \alpha + \beta = (i_2 - i_1) + (i'_2 - i'_1) = A - (i_1 + i'_1) \\ n_2/n_1 &= \sin i_1 / \sin i_2 = \sin i'_1 / \sin i'_2 = (\sin i_1 + \sin i'_1) / (\sin i_2 + \sin i'_2) \\ &= \frac{2 \sin \frac{1}{2}(i_1 + i'_1) \cos \frac{1}{2}(i_1 - i'_1)}{2 \sin \frac{1}{2}(i_2 + i'_2) \cos \frac{1}{2}(i_2 - i'_2)} \\ &= \frac{\sin \frac{1}{2}(A + \delta) \cdot \cos \frac{1}{2}(i_1 - i'_1)}{\sin \frac{1}{2}A \cdot \cos \frac{1}{2}(i_1 - i'_1 - \alpha - \beta)} & &= \frac{\sin \frac{1}{2}(A - \delta) \cdot \cos \frac{1}{2}(i_1 - i_1)}{\sin \frac{1}{2}A \cdot \cos \frac{1}{2}(i_1 - i'_1 + \alpha - \beta)}. \end{aligned}$$

Now when $i_1 > i'_1$, $\alpha > \beta$, and when $i'_1 > i_1$, $\beta > \alpha$, so that the quotients of the cosine terms have

maximum

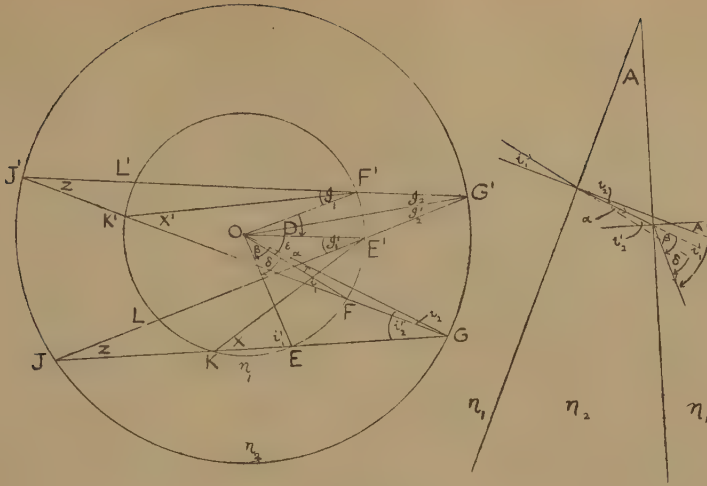
minimum.

values of unity when $i_1 = i'_1$ or $\alpha = \beta$. Thus, as n_2/n_1 and A are constants, δ has a minimum value D when $i_1 = i'_1$ given by

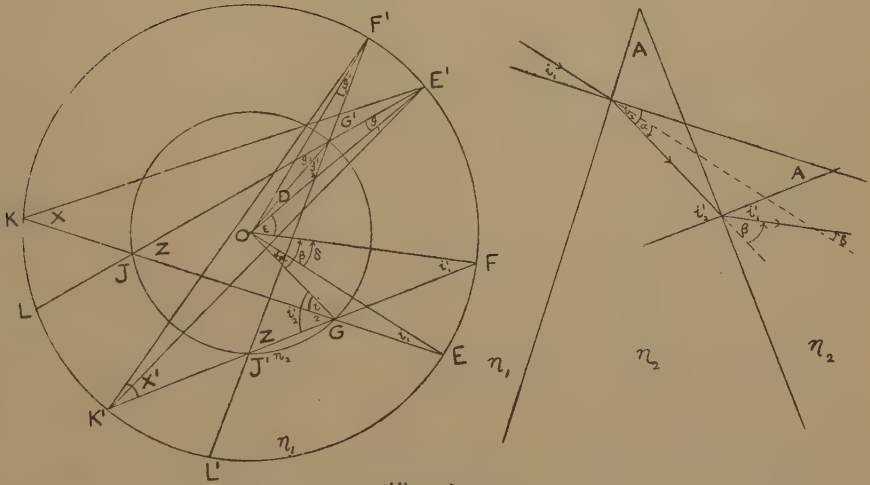
$$n_2/n_1 = \sin \frac{1}{2}(A + D) / \sin \frac{1}{2}A$$

$$n_2/n_1 = \sin \frac{1}{2}(A - D) / \sin \frac{1}{2}A$$

Fig. 3.



(i) $n_1 < n_2$.



(ii) $n_1 > n_2$.

$A < i'_2$, i. e., $A = i'_2 - i_2$.

$$\begin{aligned} \delta &= \beta - \alpha = (i'_1 - i'_2) - (i_1 - i_2) = (i'_1 - i_1) - A & \delta &= \beta - \alpha = (i'_2 - i'_1) - (i_2 - i_1) = A - (i'_1 - i_1) \\ n_2/n_1 &= \sin i_1 / \sin i_2 = \sin i'_1 / \sin i'_2 = (\sin i'_1 - \sin i_1) / (\sin i'_2 - \sin i_2) \\ &= \frac{2 \sin \frac{1}{2}(i'_1 - i_1) \cos \frac{1}{2}(i'_1 + i_1)}{2 \sin \frac{1}{2}(i'_2 - i_2) \cos \frac{1}{2}(i'_2 + i_2)} \\ &= \frac{\sin \frac{1}{2}(A + \delta) \cdot \cos \frac{1}{2}(i'_1 + i_1)}{\sin \frac{1}{2}A \cdot \cos \frac{1}{2}(i'_1 + i_1 - \alpha + \beta)} & &= \frac{\sin \frac{1}{2}(A - \delta) \cdot \cos \frac{1}{2}(i_1 + i_1)}{\sin \frac{1}{2}A \cdot \cos \frac{1}{2}(i'_1 + i_1 + \alpha + \beta)}. \end{aligned}$$

Here the quotients of the cosine terms are respectively

< 1

> 1

so that in both instances, for fixed values of A , n_1 and n_2 ,

$\delta > D$.

Alternatively, if we adopt the convention that angles of incidence and of refraction measured on opposite sides of the normal, and deviations in the opposite direction to those shown in fig. 2 (i) are to be considered negative, then the argument used for the example of fig. 2 (i) applies also to the possibilities shown in the other three figures.

It may not perhaps be out of place to remark that when a ray of light is incident nearly normally on a prism of small angle, we may write $n_2/n_1 = i_1/i_2 = i'_1/i'_2$ and deduce that $\delta = A(n_2 - n_1)/n_1$ or $A(n_1 - n_2)/n_1$ according as $n_1 < \text{or} > n_2$. Thus δ is independent of the actual value of i_1 provided it is small, and the deduction of these simple formulæ by replacing sines by angles in the formulæ $n_2/n_1 = \sin \frac{1}{2}(A \pm D)/\sin \frac{1}{2}A$, which hold only for one value of i_1 , is illogical.

III. Graphical Construction for the Deviation produced by a Prism.

The concentric circle construction may be used to investigate the deviation produced by a prism in the manner indicated in figs. 2 and 3. From O a line is drawn parallel to the direction of the incident ray, and, from the point of intersection (E) of this line with the circle of radius n_1 , another line, JEG, is drawn parallel to the normal and produced to intersect the circle of radius n_2 at G and at J. OG then represents the direction of the ray after one refraction. A line GF is then drawn from G in such a way that the angle EGF is equal to the refracting angle of the prism (A), F being the point of intersection with the circle of radius n_1 . If GF (or FG) produced cuts the n_2 circle again in J', then, angle OFJ' being the angle of emergence, the line OF and the angle EOF indicate respectively the direction of the emergent ray and the total deviation.

Regarding J and J' as fixed points, we can, by supposing G to move round the circumference of the circle JGJ', investigate the variation of say δ or i'_1 with i_1 .

(In lettering the diagrams point E (or point F) is associated with either the incident or the emergent ray in order that the argument used in section IV may apply to the four diagrams concerned.)

(i) $n_1 < n_2$.—Consideration of fig. 2 (i) makes it apparent that a ray is only transmitted through the prism when the values of angles A and i_1 lie between certain limits. Thus, when GE is a tangent to the circle of radius n_1 , $i_1 = 90^\circ$, i_2 is the critical angle (i_2^*) and the deviation produced by the first refraction is $(90^\circ - i_2^*)$. If, in addition, angle $A > 2i_2^*$, GJ' does not cut the inner circle, *i.e.* the ray becomes totally internally reflected at the second face: should angle $A = 2i_2^*$, the ray just emerges with $i'_1 = 90^\circ$: and if $2i_2^* > A > i_2^*$, both GJ and GJ' may cut the inner circle and there is a range of values of i_1 for which transmission occurs, the limits being $i_1 = 90^\circ$ and the value corresponding to $i'_1 = 90^\circ$ (GF a tangent to the inner circle).

Should the angle of the prism (JGJ') be $< i_2^*$, then as i_1 varies between 90° (GJ a tangent) and 0° (GJ passes through O) both GJ and GJ' cut

the inner circle and continue to do so as i_1 is measured on the opposite side of the normal until $i_1' = 90^\circ$ (GJ a tangent—see fig. 3 (i)).

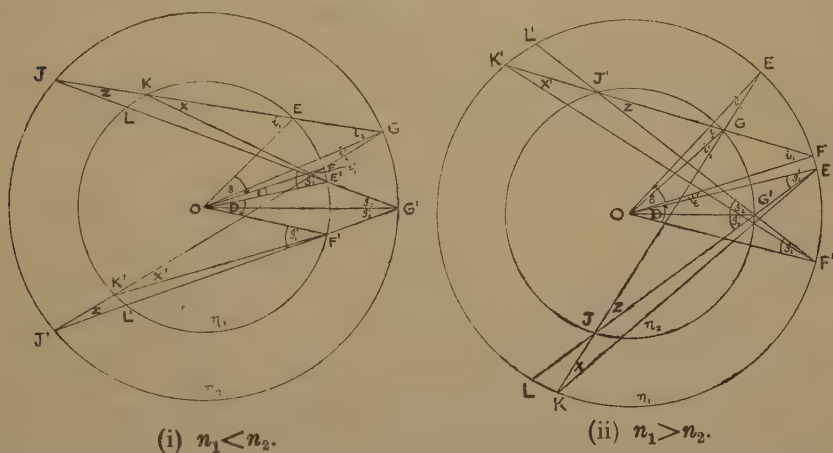
It will be observed that passage through the prism deviates the ray away from the refracting edge.

(ii) $n_1 > n_2$.—When this condition obtains there is the possibility that the incident ray may be reflected at the first surface ($i_1 >$ the critical angle, so that EGJ in figs. 2 (ii) and 3 (ii) does not cut the inner circle): total reflexion cannot of course occur at the second interface, so that any ray which penetrates the first surface is transmitted by the prism. It will, moreover, be observed that passage through the prism deviates a ray towards the refracting edge.

IV. Geometrical Proof of the Minimum Deviation Property of a Prism.

In fig. 4 the paths of rays transmitted through a prism symmetrically and unsymmetrically (angles of incidence \mathcal{J}_1 and i_1 respectively) are

Fig. 4.



traced by the concentric circle construction (see fig. 2), and lines E'K and F'K' have been added to the diagrams.

Considering the angles subtended by the arcs EE' and FF' at the centres and circumferences of the circles of radius n_1 , we have

$$\delta + \epsilon = 2X \text{ and } D + \epsilon = 2X' \text{ so that } \delta - D = 2(X - X')$$

Now, as J and J' are on the circumference of the circle JJ'G'G, angle GJG' = angle GJ'G' = Z (say), and from triangles F'J'K' and E'JK respectively we have

$$F'J'/\sin X' = F'K'/\sin Z \text{ and } E'J'/\sin X = E'K'/\sin Z.$$

Again, as K' is between the symmetrical lines E'J and F'J' and K is outside them, F'K' is nearer O than E'K, and $F'K' > E'K$. Thus $\sin X > \sin X'$, $X > X'$, and $\delta > D$, i. e. D, the angle of minimum deviation, occurs when the ray traverses the prism symmetrically.

The same result is obtained when $J'G$ and JG' intersect outside the inner circle for $n_1 < n_2$ or within the outer circle for $n_1 > n_2$, although the angles subtended by the arcs EE' and FF' are then $(\delta - \epsilon)$ and $(D - \epsilon)$.

Similarly lettered diagrams are shown in fig. 3 to demonstrate that, when the incident ray lies on the side of the normal nearer the refracting edge, δ is still $> D$.

LXXIII. *On Contact Angles.*

By ERIC J. IRONS, Ph.D., Queen Mary College, c/o King's College, Cambridge*.

[Received July 22, 1943.]

ABSTRACT.

The theory of a capillary tube method for the determination of a contact angle between a liquid and a solid is considered in detail, and the results of observations on the receding angles made by six different liquids on six different metals and by a number of liquids on glass are reported. The results indicate that, to a few parts in a thousand, the cosines of the angles are all unity.

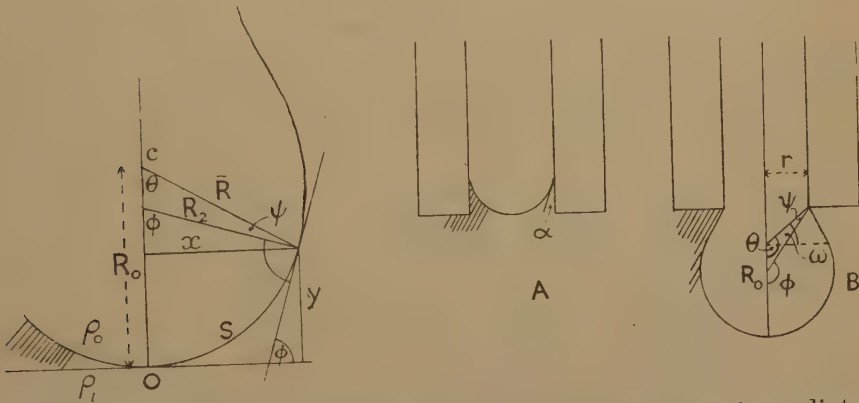
1. *Introduction.*

INTEREST in the measurement of contact angles has recently been revived by a Physical Society lecture (Ferguson, 1941), in which it was pointed out that some results that have been interpreted as showing a zero (or constant) contact angle (α) for water in contact with some metals were unwittingly obtained by a method independent of contact angle. These results (Carver and Hovorka, 1925), which purported to show the constancy of the product of capillary elevation (h) and tube radius (r) whatever the metal, were obtained by means of a modified form of the apparatus originally designed by Bigelow and Hunter (1911) to investigate the suggestion that, for any one liquid, the product rh depended upon the material of the tube. Bigelow and Hunter, unlike Carver and Hovorka, inferred that rh was not constant and reported, among others, average values of 15.15, 15.11, 14.75, and 14.69 mm.² for water in tubes of zinc, copper, silver and glass respectively. They were inclined to interpret these variations as due to some electrical effect; and indeed their results hardly admit of complete interpretation in terms of contact angle when it is remembered that, if we assume $\alpha=0$ and take the value of the surface tension of water as 73 dyne. cm.⁻¹, rh has the value 14.88 mm.², which is less than that noted for zinc and copper. During a discussion on the state of these matters with Prof. Ferguson, it occurred to the author that values for the cosines of the contact angles made by

* Communicated by the Author.

liquids with metals might be determined very conveniently by a capillary tube method involving the comparison of two pressures. The method then contemplated, and subsequently used to obtain the results herein reported, consists of placing a capillary tube of the metal vertically in a bath of the liquid and finding (A) the pressure required to maintain the liquid meniscus at the lower end of the capillary (Ferguson and Dowson, 1921), and then (B) the maximum pressure occurring as a bubble of air is detached (*cf.* Jäger, 1891). Experiment shows that these two pressures do not differ greatly, or, what is equivalent, that the cosines to be measured are approximately unity: it is, in consequence, necessary to consider carefully the magnitudes of certain correcting terms if the existence of small but finite contact angles are to be established. A treatment convenient for this purpose is provided by the work of Verschaffelt (1918), but, possibly because it only forms part of an investigation or possibly because the analysis is long, the steps in the

Fig. 1.



argument and the approximations made at each stage are not immediately obvious. It has therefore been felt desirable to place on record a more detailed analysis, and this is included in the next section.

2. Theory of the Method.

The curved line in fig. 1 represents a liquid boundary which may be formed either in part within a capillary tube or in its entirety at the bottom of such a tube. The equation to the section, considered as part of a surface of revolution, is

$$\gamma(1/R_1 + 1/R_2) = g(\rho_1 - \rho_0)(h + y), \quad \dots \dots (1)$$

where γ , ρ , and R represent respectively surface tension, density, and radius of curvature (R_1 being the radius in the plane of the paper, R_2 the radius in a plane perpendicular to the paper, and R_0 the value of both these radii at O); g the acceleration due to gravity; and $h = \gamma g^{-1}(\rho_1 - \rho_0)^{-1}(2/R_0)$, the head corresponding to the pressure difference across the boundary at O .

Now, in terms of the symbols marked on the figure, $1/R_1 = (d\phi/ds) = (d\phi/dx)(dx/ds) = (d\phi/dx) \cos \phi$ and $1/R_2 = (\sin \phi)/x$, so (1) may be written as

$$\frac{d\phi}{dx} \cdot \cos \phi + \frac{\sin \phi}{x} = \frac{g(\rho_1 - \rho_0)}{\gamma} (h + y) = \frac{h + y}{a^2} \text{ (say)}, \quad \dots \quad (2)$$

which has as solution

$$x \cdot \sin \phi = \frac{1}{2}(h/a^2)x^2 + (1/a^2) \int_0^x xy \cdot dx. \quad \dots \quad (2')$$

By (2') we note that, if y is negligible compared with h , $x = (2a^2/h) \sin \phi = R_0 \sin \phi$, and the section is circular. For our present purpose, however, we may only regard y as small compared with h and the integral must be evaluated. To this end, and to avoid discontinuities when $\theta = \pi/2$, (2) is expressed in polar co-ordinates with C , the centre of curvature of the bottom of the curve, as origin, by putting $x = \bar{R} \sin \theta$ and $y = R_0 - \bar{R} \cos \theta$. Thus, writing $\dot{\bar{R}}$ for $(d\bar{R}/d\theta)$ and remembering that

$$\tan \phi = \frac{dy}{dx} = \frac{d}{d\theta}(R_0 - \bar{R} \cos \theta) \bigg/ \frac{d}{d\theta}(\bar{R} \sin \theta),$$

it is readily shown that

$$\cos \phi = \frac{\bar{R} \cos \theta + \dot{\bar{R}} \sin \theta}{\sqrt{(\bar{R}^2 + \dot{\bar{R}}^2)}}, \quad \sin \phi = \frac{\bar{R} \sin \theta - \dot{\bar{R}} \cos \theta}{\sqrt{(\bar{R}^2 + \dot{\bar{R}}^2)}},$$

and
$$\frac{d\phi}{dx} = \frac{d\phi/d\theta}{dx/d\theta} = \frac{\bar{R}^2 - \bar{R}\ddot{\bar{R}} + 2\dot{\bar{R}}^2}{(\bar{R} \cos \theta + \dot{\bar{R}} \sin \theta)(\bar{R}^2 + \dot{\bar{R}}^2)},$$

so that, substituting in (2),

$$\frac{\bar{R}^2 - \bar{R}\ddot{\bar{R}} + 2\dot{\bar{R}}^2}{(\bar{R}^2 + \dot{\bar{R}}^2)^{3/2}} + \frac{\bar{R} \sin \theta - \dot{\bar{R}} \cos \theta}{\bar{R} \sin \theta (\bar{R}^2 + \dot{\bar{R}}^2)^{1/2}} = \frac{2}{R_0} + \frac{R_0 - \bar{R} \cos \theta}{a^2}. \quad \dots \quad (3)$$

As this equation cannot be integrated directly, it is assumed that the deviation of the section from circularity may be expressed by writing $\bar{R} = R_0(1 - \tau)$, and therefore $\dot{\bar{R}} = -R_0\dot{\tau}$, and $\ddot{\bar{R}} = -R_0\ddot{\tau}$, as τ is a function of θ . Neglecting τ^2 and $\dot{\tau}^2$ in comparison with unity, (3) now becomes

$$2\tau \sin \theta + \dot{\tau} \cos \theta + \ddot{\tau} \sin \theta = (R_0^2/a^2) \sin \theta (1 - \bar{1} - \tau \cos \theta) \\ \div (R_0^2/a^2) \sin \theta (1 - \cos \theta),$$

since in practice R_0 is small in comparison with $h (= 2a^2/R_0)$, i. e., R_0^2/a^2 is small compared with unity, and this applies also to τ .

Multiplying throughout by the integrating factor $\cos \theta$, and integrating

$$(\tau \sin \theta + \dot{\tau} \cos \theta) \sin \theta = (R_0^2/a^2) \left(\frac{1}{2} \sin^2 \theta + \frac{1}{3} \cos^3 \theta - \frac{1}{3} \right),$$

since $\theta = 0$ when $\tau = 0$. Dividing throughout by $\sin \theta \cos^2 \theta$ and integrating again gives

$$\tau \sec \theta = (R_0^2/a^2) \left\{ \left[\frac{1}{6 \cos \theta} + \frac{1}{3} \log (\sin \theta) - \frac{1}{3} \log (\tan \theta/2) \right] - \left[\frac{1}{6} + \frac{1}{3} \log 2 \right] \right\},$$

having regard to the same limits as before. Thus

$$\tau = \frac{1}{6} (R_0^2/a^2) \left[(1 - \cos \theta) + 2 \cos \theta \cdot \log \left(\frac{1 + \cos \theta}{2} \right) \right], \quad (4)$$

and, differentiating with respect to θ ,

$$\dot{\tau} = \frac{1}{6} (R_0^2/a^2) \sin \theta \left[\frac{1 - \cos \theta}{1 + \cos \theta} - 2 \log \left(\frac{1 + \cos \theta}{2} \right) \right], \quad (5)$$

and both (4) and (5) remain valid over the range of values of θ with which we are concerned.

Moreover, as

$$\sin \phi = \frac{\bar{R} \sin \theta - \dot{\bar{R}} \cos \theta}{\sqrt{(\bar{R}^2 + \dot{\bar{R}}^2)}} = \frac{R_0(1-\tau) \sin \theta + R_0 \dot{\tau} \cos \theta}{R_0(1-\tau)[1 + \dot{\tau}^2/(1-\tau)^2]^{1/2}} = \sin \theta + \dot{\tau} \cos \theta$$

to the same approximation as before, and as $\phi = \theta + \psi$, therefore

$$\psi = \dot{\tau}. \quad (6)$$

It is now possible, by substituting the approximate values of x and y in the small term of (2'), to obtain a more exact solution of (2). Thus

$$\begin{aligned} \int_0^x xy \cdot dx &= \int_0^\theta (R_0 \cdot \overline{1-\tau} \cdot \sin \theta) (R_0 - R_0 \cdot \overline{1-\tau} \cdot \cos \theta) R_0 (\overline{1-\tau} \cos \theta - \dot{\tau} \sin \theta) d\theta \\ &= R_0^3 \int_0^\theta [(\sin \theta \cos \theta - \sin \theta \cos^2 \theta) + \tau(-2 \sin \theta \cos \theta + 3 \sin \theta \cos^2 \theta) \\ &\quad + \dot{\tau}(-\sin^2 \theta + \sin^2 \theta \cos \theta)] d\theta, \text{ neglecting terms in } \tau^2 \text{ and } \tau \dot{\tau}, \\ &= \frac{1}{6} R_0^3 (1 - \cos \theta)^2 (1 + 2 \cos \theta) + \frac{1}{6} (R_0^5/a^2) (1 - \cos \theta)^2 \cos^2 \theta \\ &\quad + \frac{1}{6} (R_0^5/a^2) \sin^2 \theta (1 - 2 \cos \theta + 2 \cos^2 \theta) \cdot \log \left(\frac{1 + \cos \theta}{2} \right), \quad (7) \end{aligned}$$

on substituting for τ and $\dot{\tau}$, integrating and collecting terms. Remembering the definition of h we may write (2') in the form

$$x = R_0 \sin \phi - \frac{R_0}{xa^2} \int_0^x xy \cdot dx = R_0 \sin \phi - [(1+\tau)/(a^2 \sin \theta)] \int_0^x xy \cdot dx,$$

putting $(1+\tau)$ for $1/(1-\tau)$.

Now, on substituting in this expression the values of τ and of the integral from equations (4) and (7) respectively, the second term is obtained as a function of θ ; but if, by virtue of the same argument as that leading to equation (6), $(\sin \phi - \dot{\tau} \cos \phi)$ and $(\cos \phi + \dot{\tau} \sin \phi)$ may be written for $\sin \theta$ and $\cos \theta$ respectively, and, as $\dot{\tau}$ is a small quantity, ϕ may replace θ in the terms which involve $\dot{\tau}$ (actually those containing the factor R_0^5/a^4); then, after some reduction in which terms involving higher powers of R_0 are neglected, we obtain

$$x = R_0 \cdot \sin \phi - \frac{1}{6} \frac{R_0^3}{a^2} \cdot \frac{\sin \phi (1 - \cos \phi) (1 + 2 \cos \phi)}{1 + \cos \phi}$$

$$-\frac{1}{36} \frac{R_0^5}{a^4} \cdot \frac{\sin \phi (1 - \cos \phi) (1 - 3 \cos \phi + 6 \cos^2 \phi + 8 \cos^3 \phi)}{(1 + \cos \phi)^2} \\ - \frac{1}{6} \frac{R_0^5}{a^4} \cdot \sin \phi \cdot \log \left(\frac{1 + \cos \phi}{2} \right), \quad \dots \quad (8)$$

which is the fundamental equation from which are derived the formulæ appropriate to experiments of the types (A) and (B).

(A.) Supposing the lower portion of the curve of fig. 1 to represent the surface, at the bottom of a capillary tube of radius r , of a liquid which makes a contact angle α with the material of the tube (fig. 1 A), we have, on substituting $x=r$, $\phi=\pi/2-\alpha$, and $R_0=2a^2/h$ in (8),

$$r = \frac{2a^2}{h} \cos \alpha \left[1 - \frac{2}{3} \frac{a^2}{h^2} \cdot \frac{(1 - \sin \alpha)(1 + 2 \sin \alpha)}{1 + \sin \alpha} \right. \\ \left. - \frac{4}{9} \frac{a^4}{h^4} \cdot \frac{(1 - \sin \alpha)(1 - 3 \sin \alpha + 6 \sin^2 \alpha + 8 \sin^3 \alpha)}{(1 + \sin \alpha)^2} \right. \\ \left. - \frac{8}{3} \frac{a^4}{h^4} \cdot \log \left(\frac{1 + \sin \alpha}{2} \right) \right].$$

This equation is satisfied by

$$\frac{2a^2}{h} = r \sec \alpha \left[1 + \frac{1}{3} \frac{r}{h} \cdot \sec^3 \alpha (1 - \sin \alpha)^2 (1 + 2 \sin \alpha) \right. \\ \left. + \frac{1}{3} \frac{r^2}{h^2} \cdot \sec^6 \alpha (1 - \sin \alpha)^3 (1 + \sin \alpha + 2 \sin^2 \alpha) \right. \\ \left. + \frac{2}{3} \frac{r^2}{h^2} \cdot \sec^2 \alpha \cdot \log \left(\frac{1 + \sin \alpha}{2} \right) \right], \quad \dots \quad (9)$$

as may be verified by substitution; and we may note in passing that, if $\alpha=0$,

$$a^2 = \frac{hr}{2} \left[1 + \frac{1}{3} \frac{r}{h} - \frac{1}{3} \frac{r^2}{h^2} (2 \log 2 - 1) \right].$$

(B.) Let us next suppose that, by the application of pressure, the liquid surface is ejected from the end of the tube to the position shown in fig. 1 B and that a bubble of air is about to be detached: it is required to determine the maximum value of the pressure which occurs during this process. Putting $\theta=\pi/2+\omega$, assuming ω to be small, and substituting in (4) and (5) gives

$$\tau = \frac{1}{6} R_0^2/a^2 \quad \text{and} \quad \dot{\tau} = \frac{1}{6} (R_0^2/a^2) (1 + 2 \log 2),$$

while, from a consideration of conditions at the bottom of the capillary tube, we have

$$(i) \quad x=r=R_0(1-\tau) \sin \theta = R_0(1-\tau)(1-\frac{1}{2}\omega^2),$$

whence, differentiating with respect to θ (or ω) and neglecting terms in $\tau\dot{\tau}$ and ω^3 ,

$$\dot{R}_0 = R_0 \left[\omega + \frac{1}{6} (R_0^2/a^2) (1 + 2 \log 2) \right].$$

$$(ii) \quad h+y=2a^2/R_0+R_0-\bar{R} \cos \theta=2a^2/R_0+R_0+R_0\omega.$$

Differentiating with respect to ω and equating the differential coefficient to zero to find the condition for maximum pressure gives

$$\dot{R}_0=R_0^3/2a^2$$

on neglecting terms containing powers of $R_0^2/2a^2$ higher than the first.

Equating these values of \dot{R}_0 shows that

$$\omega = \frac{1}{3} (R_0^2/a^2)(1-\log 2),$$

and is positive, *i. e.*, the maximum pressure occurs when the bubble has passed the hemispherical stage in its formation.

Again, $\phi=\theta+\psi=\pi/2+\omega+\psi$ and $\psi=\tau=\frac{1}{6}(R_0^2/a^2)(1+2\log 2)$, so

$$\sin \phi = \left(1 - \frac{1}{2}\omega^2\right) \left(1 - \frac{1}{2}\psi^2\right) - \omega\psi = \left(1 - \frac{1}{8}R_0^2/a^2\right) \text{ and}$$

$$\cos \phi = -\left[\omega\left(1 - \frac{1}{2}\psi^2\right) + \left(1 - \frac{1}{2}\omega^2\right)\psi\right] = -\frac{1}{2}R_0^2/a^2.$$

Substituting $x=r$ and the values of $\sin \phi$ and $\cos \phi$ in (8) we have, neglecting terms in R_0^6/a^6 ,

$$r=R_0\left[1-\frac{1}{6}\frac{R_0^2}{a^2}-\frac{1}{72}\frac{R_0^4}{a^4}(11-12\log 2)\right],$$

$$\text{whence,} \quad R_0=r\left[1+\frac{1}{6}\frac{r^2}{a^2}-\frac{1}{72}\frac{r^4}{a^4}(12\log 2-17)\right],$$

as may be verified by substitution.

Putting H for the maximum value of $(h+y)$ and substituting the values found for ω and R_0 in the expression $2a^2/R_0+R_0(1+\omega)$ gives

$$H=\frac{2a^2}{r}+\frac{2}{3}r+\frac{1}{12}\frac{r^3}{a^2},$$

$$\text{whence} \quad \frac{2a^2}{H}=r-\frac{2}{3}\frac{r^2}{H}-\frac{1}{12}\frac{r^4}{H\cdot a^2}=r\left(1-\frac{2}{3}\frac{r}{H}-\frac{1}{6}\frac{r^2}{H^2}\right), \quad (10)$$

putting $rH/2$ for a^2 in the third term.

Now equation (9) relates the angle of contact to the pressure difference between the two sides of the liquid boundary when its centre is situated at the level of the bottom of the capillary tube, and equation (10), which does not involve α , gives the maximum pressure that exists at the same level during the detachment of an air bubble; it follows that from measurements of these two pressures it is possible to deduce the value of α .

Dividing one equation by the other yields

$$\cos \alpha = (h/H)\left(1+\frac{1}{3}r/h\right)/\left(1-\frac{2}{3}r/H\right) = \left(h+\frac{1}{3}r\right)/\left(H-\frac{2}{3}r\right), \quad (11)$$

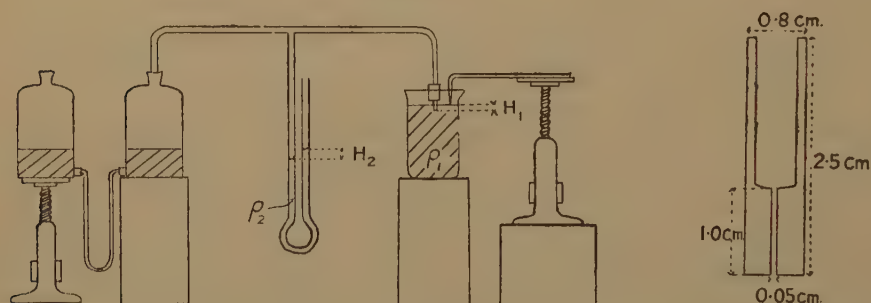
if the terms in r^2/h^2 and r^2/H^2 are negligible, as it will be shown is usual, and $\sec^3 \alpha (1 - \sin \alpha)^2 (1 + 2 \sin \alpha)$ may be equated to unity. Should $\cos \alpha$, as determined by this equation, differ appreciably from unity the approximate value of the coefficient of $\frac{1}{3}r/h$ may be calculated and a more exact value of α found.

It should be observed that r , the tube radius, only occurs in the small terms of the expression for $\cos \alpha$, so that one of the advantages of this method of measuring α is that it does not involve a very accurate measurement of r . Other advantages are that it is possible to make the required setting for the determination of h with the meniscus either advancing or retreating in the tube, and that determinations of h and of H may be made in rapid succession with newly created liquid surfaces.

3. Experimental.

The method of measurement may be inferred from fig. 2. By means of the elevating table on the left-hand side of the apparatus the pressure inside the capillary tube could be varied at will and measured in terms

Fig. 2.



of the height H_2 on the water manometer. To determine H and h (see last section) it was also necessary to find H_1 (see fig. 2), and to this end the surface of the liquid was located by a fine glass point at the conclusion of each experiment, the required distance being measured subsequently by a travelling microscope. For experiments of type A the meniscus in the capillary tube was viewed through a microscope with its axis at a slight angle to the horizontal, and the pressure was adjusted until the bottom of the meniscus appeared to bisect the foreshortened opening of the cylindrical hole in the metal: values of H_1 were recorded when the meniscus was adjusted from positions both within the tube (\downarrow) and from just outside the tube (\uparrow). To determine the value of the maximum pressure in an experiment of type B, the cross-wires of a microscope were adjusted by a few trials so as to be opposite the turning position of the manometer meniscus when a bubble was released.

The metal tubes employed were supplied and first drilled by Messrs. Johnson, Matthey and Co., Ltd., and were of the dimensions shown in fig. 2: they were subsequently trimmed where necessary so as to be

free from any burr or flare. Mean values of tube diameters were estimated in terms of the divisions of the eyepiece scale of a Swift microscope, which was calibrated by a stage micrometer.

Glassware was cleaned by chrome-sulphuric acid followed by tap water, the beaker containing the capillary being carefully dried and washed out with the liquid under test. In experiments with a glass capillary any remaining tap water was removed by washing with alcohol followed by acetone or by ether which was allowed to evaporate. Some of the liquid to be tested was then drawn up and down the capillary and finally ejected; this was repeated a number of times, the actual measurements being made with a fresh sample. The metal tubes were kept in benzene and washed with a fresh supply, which was allowed to evaporate before the capillary was finally treated with the liquid under test. The metal tube was attached to the main apparatus by a rubber sleeve.

The liquids employed were purified by repeated distillation with the exceptions of acetic acid, which was recrystallized, and pinene which was supplied by Messrs. Harrington and Co., Ltd., and untreated: the ether received some preliminary drying.

To conclude the section I have pleasure in recording that this work was made possible by the assistance of Mr. H. Saffrey, the Queen Mary College Laboratory Steward.

4. Results and Discussion.

The method of deducing the value of $\cos \alpha$ is illustrated by the example given in Table I.

TABLE I.

Tube :—Aluminium. Liquid :—Water. Temperature :—16°.

Readings of microscope focused on manometer meniscus :—

Experiment A.	Experiment B.	Zero.
7.268 ↓	7.289	4.054 cm.
7.268 ↑		
7.267 ↓	7.288	∴ H_{2A} (see fig. 2)
7.267 ↓	7.287	$= 2 \times 3.214 = 6.42(8) \text{ cm.}$
7.269 ↑		∴ H_{2B}
Mean 7.268 cm.	7.288 cm.	$= 2 \times 3.234 = 6.46(8) \text{ cm.}$

Reading of microscope focused on :—

Surface of liquid = 6.254 cm.; Bottom of capillary = 5.715 cm.;

∴ H_1 (see fig. 2) = 0.539 cm.

Relative density of liquid (normally determined by specific gravity bottle) = ρ_1/ρ_2 (see fig. 2) = 1.000

(ρ_0 , the density of air, is neglected.)

Mean diameter of capillary tube = 0.050(8) cm.,

Now $h + \frac{1}{3}r = H_{2A} \cdot \rho_2 / \rho_1 - (H_1 - \frac{1}{3}r)$ (cf. eqn. (9))
 and $H - \frac{2}{3}r = H_{2B} \cdot \rho_2 / \rho_1 - (H_1 + \frac{2}{3}r)$ (cf. eqn. (10)),
 so that by eqn. (11),

$$\cos \alpha = \frac{H_{2A} - (\rho_1 / \rho_2)(H_1 - \frac{1}{3}r)}{H_{2B} - (\rho_1 / \rho_2)(H_1 + \frac{2}{3}r)} = \frac{6.428 - (0.539 - 0.008)}{6.468 - (0.539 + 0.017)} = \frac{5.897}{5.912} = 0.997$$

and, by eqns. (2) and (10),

$$\gamma = g\rho_1 a^2 = (g\rho_1 r/2)(H - \frac{2}{3}r) = (g\rho_2 r/2)\{H_{2B} - (\rho_1 / \rho_2)(H_1 + \frac{2}{3}r)\} \\ = 981 \times 0.999 \times (0.0508/4) \times 5.912 = 73.6 \text{ dyne cm.}^{-1}.$$

Note that, as $r/h \simeq r/H \simeq 0.025/6.0 \simeq 0.004$, it is permissible to neglect the terms in $(r/h)^2$ and $(r/H)^2$ (see eqn. (11)).

In Table II. are summarized some results obtained in preliminary experiments with a glass capillary. These experiments were undertaken

TABLE II.
Values obtained using a glass capillary.

Water.	Benzene.	Acetone.	Carbon tetrachloride.	Aniline.
1.000	1.002	0.998	1.003	1.001
73.6	28.8	23.7	27.7	44.0
18	18	19	15	21

Pinene.	Acetic acid.	Cyclohexanone.	Ether.	Paraffin (medicinal).
1.000	1.001	1.001	0.999	0.999
27.2	27.8	34.6	17.0	31.8
17	21	19	20	19

not only to investigate the possibilities of the method but also, in view of the finite contact angles that have been reported, to verify the more modern belief that, for most liquids, the angle has a zero value (International Critical Tables, 1928). The numbers shown in this table and in Table III. are, taken in order, the values of the cosine of the contact angle, the surface tension (dyne cm.⁻¹), and the temperature (°C.). The surface tension values were calculated from the results of the experiments of type B as a check on the readings, although of course the actual figures reported depend upon the tube radii, which, for the metals, could not be determined with precision.

The values for the cosines which are greater than unity indicate that the accuracy attained may not exceed one part in two hundred, so that it is only safe to conclude that the angles measured lie between 0° and 6° (cos⁻¹ 0.995): a temptation to regard them as zero or very small has some justification in the case of glass, where, possibly on account

of the regularity of the capillary, the values of the cosines are not so scattered.

It should perhaps be emphasized that the zero (or small) contact angles assigned to the liquid-metal combinations as a result of these measurements are for surfaces which have been previously covered by the liquid, *i. e.* the angles are essentially *receding angles*. For the determination of advancing angles it would be necessary to adopt an alternative method of cleaning or, for example, to follow the procedure of Bartell and Cardwell (1942), who have formed silver and gold surfaces *in vacuo* from the vapour phase. These authors have shown that water

TABLE III.

	Water.	Benzene.	Acetone.	Carbon tetra- chloride.	Aniline.	Pinene.
Aluminium	0.997 73.7 16	0.993 28.9 19	0.997 24.7 19	0.999 27.5 17	0.995 43.7 20	0.997 27.3 15
Iron	1.001 73.4 18	0.999 29.0 19	1.002 25.1 18	1.006 27.6 16	0.999 43.9 21	1.002 27.1 21
Nickel	0.994 72.5 16	0.995 28.6 18	0.998 24.0 16	1.000 27.5 17	0.994 43.6 18	0.995 27.8 18
Copper	0.998 73.8 15	1.005 29.5 14	1.005 24.3 15	1.004 26.4 22	1.003 42.4 21	1.006 26.8 24
Silver	1.001 73.0 16	1.003 28.8 18	1.005 23.6 17	1.007 27.1 18	1.004 43.4 20	1.004 26.8 18
Platinum	0.998 72.9 18	1.004 27.8 19	1.005 23.6 19	1.003 26.8 20	1.002 43.3 21	1.001 27.1 21

It should be explained that values of $\cos \alpha$ differing considerably from unity were occasionally obtained, but, as they did not recur after further cleaning and reboiling, they were rejected.

makes similar and reproducible contact angles when advancing over such surfaces, and have also always observed a zero angle when a water drop was formed on a silver surface, allowed to stand for a few minutes, and then caused to recede. This they suggest may be due to the immobilization of the periphery of the drop by water which, by capillary action, has become retained in sub-microscopic surface pores or cracks, so that as the bulk of the liquid disappears the angle necessarily approaches zero. The values herein recorded serve, at least, to confirm their observations.

As a result of the present work we may conclude that the method affords a convenient means of estimating a receding contact angle and, given a

regular capillary, that it is capable of reasonable accuracy; we may also infer that the product rh (Section 1) would be sensibly independent of the material of a wetted tube, *i. e.*, the condition and immediate past history of a surface are, in this connexion, factors of greater importance than its material.

This investigation was carried out in the Cavendish Laboratory and my thanks are due, and are tendered, to Prof. Sir Lawrence Bragg for providing the necessary accommodation.

References.

- Bartell and Cardwell, *J. Amer. Chem. Soc.* lxiv. p. 494 (1942).
 Bigelow and Hunter, *J. Phys. Chem.* xv. p. 367 (1911).
 Carver and Hovorka, *J. Amer. Chem. Soc.* xlvii. p. 1325 (1925).
 Ferguson, *Proc. Phys. Soc.* liii. p. 554 (1941).
 Ferguson and Dowson, *Trans. Faraday Soc.* xvii. p. 384 (1921).
 International Crit. Tables, iv. p. 434 (1928).
 Jäger, *K. Akad. Wiss. Wien, Ber.* (2 a), c. p. 245 (1891).
 Verschaffelt, *Proc. K. Ned. Akad. Wet.* xxi. pp. 357 and 366 (1918).

LXXIV. *The Problem of the Keyboard Instrument.*—II.

By LL. S. LLOYD *.

[Received June 7, 1943.]

IN an earlier paper † I indicated the nature of the contribution to musical acoustics made, in the middle of the eighteenth century, by Dr. Robert Smith in his 'Harmonics.' No mention of the noteworthy achievements of this distinguished musical mathematician would give an adequate idea of his work unless it included some account of the objective and nature of the temperament he devised for the harpsichord. This was intended to replace mean-tone temperament and to provide a better approximation to the musical scale embodied, in various keys, in music as he knew it, and particularly that of Corelli. It was to be preferred to equal temperament, which even in his day commanded some support. The claims of Dr. Smith's temperament rested on careful observation and practical trial. In this it provided an example that might well have been followed by subsequent theorists, and supplied a corrective of those arithmetical speculations which too often misled those of them who indulged in "arm-chair" theory.

An example of Dr. Smith's appeal to experiment for the verification of his conclusions about beats was given in the earlier paper. A striking example of his careful observation may be found in his discussion of the differing dissonance of different discords. This he advanced in the course

* Communicated by Sir Edward Appleton, F.R.S.

† *Phil. Mag.* (ser. 7) xxxiv. pp. 472-479 (1943).

of a reasoned refutation of Sauveur's theories, on page 96 of his work : "As Mr. Sauveur appeals to numbers let us see what evidence they produce." He proceeds to point out that, according to Sauveur's theory, "the major fourth" and "the minor fifth" would beat more slowly than an interval of a whole tone or the major or minor sevenths, all these intervals being discords. The use of the terms "major" and "minor" in this sentence implies that he was thinking of *diatonic* discords. On the white notes of the keyboard there is one fourth which is greater (major) than any others—that between F and the B above it. This is the tritone with which students of counterpoint make an early acquaintance, for it needed careful handling in music for unaccompanied voices : "*Tritonus diabolus est.*" Similarly, there is one fifth which is less (minor) than all the others—that between B and the F above it. To-day we call it the imperfect fifth. In passing it may be observed that, if measured upwards from C, these diatonic intervals become CF \sharp , the tritone of the key of G, and CG \flat , the imperfect fifth of the key of D \flat . In relation to the key of C, however, they would be *chromatic* intervals, the augmented fourth and the diminished fifth ; and this distinction of name denotes a different musical meaning. To mark this difference of meaning is not unimportant, because it was ignored by those nineteenth-century theorists who imagined that musical intervals ought to be determined by theoretical frequency-ratios instead of by the sensitive ears and musical instinct of composers and skilled artists.

The frequency-ratio of the major tone is 9 : 8 while, in theory, that of the tritone is 45 : 32. Dr. Smith pointed out that, according to Sauveur, it followed from these ratios that the tritone would beat about four or five times slower than the major tone. He observed that this was not true. "But in sounding the latter discords [the major fourth and the minor fifth] upon an Organ, Harpsichord or Violoncello, even at a low pitch, I find their beats are so quick that I cannot count them ; or rather they do not beat at all, like imperfect [mistuned] consonances, but only flutter at a slower or quicker rate according to the pitch of the sounds." Examine this observation in the light of Helmholtz's theory of dissonance. In a tritone we have to go to the inaudible forty-fifth partial tone of the lower note and the inaudible thirty-second partial tone of the upper note to reach a theoretical unison between the partial tones that would beat if the interval were mistuned. That is why the tritone is so lacking in definition, and readily changes into an imperfect fifth in an enharmonic modulation. Whether it is tuned with theoretical accuracy or not, there is rapid beating between various partial tones ; the dissonance is due to what Dr. Smith called "a fluttering roughness." A similar lack of definition is found in many chromatic intervals, such as the augmented fourth and diminished fifth. That is why, under the influence of the melodic line, G \flat may often be played below F \sharp on the violin, or C \sharp above D \flat *.

* *Ibid.* p. 475.

This example of Dr. Smith's acute powers of observation is particularly apposite, for, as we shall see, in devising his temperament he was concerned to distribute equally, between important concords, the dissonance occasioned by any tempering of the intervals. By way of contrast we may turn to the efforts of theorists of the nineteenth century. The methods of tuning keyboard instruments which they devised fall into two classes, of which mean-tone temperament and equal temperament are the representative types. Tunings in the first of these classes were intended to provide intervals as consonant as possible for use with a more or less extensive range of keys, though not for all keys. The inadequacy of mean-tone temperament, admirable as it had been for modal music, led theorists to devise tunings which gave true intervals when used for later music. One of the earliest of these efforts was that of Perronet Thompson, of which an interesting recollection is to be found in a contribution made by Dr. Southgate to the discussion of a paper on temperaments read before the Musical Association on January 15, 1902 * :—

“One instrument has been made which, perhaps, none of you but myself has heard ; it was planned by Col. Perronet Thompson, a notable mathematician and enthusiastic musician, who wrote a remarkable book showing that there should be seventy-two available sounds to the octave, but he was satisfied with forty for practical use. An organ so constructed was erected in the Welsh Chapel, Aldersgate Street, and was played by a Miss Stafford Northcote, a blind girl. I have a distinct remembrance of hearing her play Mendelssohn's ‘Wedding March,’ and I thought I had never heard such exquisite chords, and they sounded singularly rich and perfect. On the ingenious keyboard, besides the ordinary black and white keys, there were various devices of hooks, points, quadrils, buttons, sticking up for the fresh intervals introduced, and this gifted girl managed to play on it. I very much doubt whether anyone else would be able to do the same.”

A diagram of Perronet Thompson's keyboard will be found in ‘The Oxford Companion to Music’ (Dr. Percy Scholes) on Plate 90. It certainly does not encourage any organist to emulate Miss Northcote. There is an account of Perronet Thompson's organ in the translator's Appendix to Helmholtz's ‘Sensations of Tone,’ from which we learn that it was tuned with just (true) intervals and was available for 21 keys. From the same source we learn that in his theoretical work, though not for his organ, Perronet Thompson made use of Mercator's cycle of 53 †.

A more manageable keyboard was provided at a later date by Colin Brown for his Voice-Harmonium, which also gave just intervals. This, too, is described by Ellis in the Appendix referred to, but I mention it

* Proc. Mus. Assoc. xxviii. p. 63 (1902).

† Musicians are indebted to Mr. Dufton for correcting, by means of contemporary documentary evidence, Ellis's assumption that this temperament was devised only for Pythagorean intonation. Phil. Mag. (ser. 7) vol. xxxiii. p. 156 (1942).

because the last surviving instrument was presented, just before the war, by Colin Brown's daughter, Mrs. Sloan (an old lady who has since died), to the Science Museum, where it will be available for examination after the war.

Another example of an instrument designed to give true intervals in a considerable range of keys was Helmholtz's two-manual harmonium. In theory it was tuned in a temperament that gave a true major third and a perfect fifth flattened by only one-eighth of the error of the fifth of equal temperament. In practice the fifth had to be tuned true, for its theoretical temperament eluded the tuner's ear. Its object, like that of Colin Brown's instrument, was primarily experimental. It was not easy to finger.

The second class of tunings was intended to allow of completely free modulation, for which purpose use was made of a closed cycle of notes. In his article on Sound, in the 'Dictionary of Applied Physics' (1923), Barton writes: "The way to find such practical solutions or temperaments is to seek some smaller number of steps in the octave such that certain exact numbers of them will give for each note of the scale a tolerable approximation to the concord sought." It is idle, however, to stop there. Three more things remain to be done before the new temperament is worthy of consideration: these are, to construct an instrument which will give the desired number of tones, so that the ear can judge of the success of the attempt, to devise a keyboard to play it with, and to discover practicable means of tuning it correctly. All three steps were duly taken by Dr. Robert Smith. Barton gives four examples of tunings which give closed cycles of this kind.

First, Bosanquet's temperament with 53 tones to the octave. This was Mercator's temperament. It was described by Helmholtz, in a paragraph he added to the fourth edition of 'Tonempfindungen,' and in a new Appendix, No. XIX, which gives a diagram of Bosanquet's generalized keyboard, with 84 "keys" to the octave, to facilitate fingering. Readers may consult this work, or the second edition of 'Sensations of Tone,' for details of this keyboard and the harmonium and organ for which it was used. The tuning of these instruments was complicated, and apparently a trifle uncertain; but at least, like Robert Smith, Bosanquet carried his suggestion through all the necessary stages. More interesting to hear, to-day, than his use of Mercator's cycle would be his use, on his own organ, of the generalized keyboard for 36 notes of mean-tone temperament in the octave, brought into separate action by a stop.

Second, Huygen's cycle of 31 notes to the octave, which gave a close approximation to mean-tone tuning, making it available for all keys. Dr. Smith's 'Harmonics' suggests, to the modern reader, that some tuners had tried this temperament on an organ with the familiar manual.

Third, Woolhouse's cycle of 19 notes, which has little to commend it. It used 3 of the 19 divisions of the octave for the whole tone, 2 divisions for the limma representing a diatonic semitone, and 1 division to separate

adjacent sharps and flats. I mention this cycle because it was known to Robert Smith, as we shall see.

Fourth, equal temperament, which divides the octave into 12 equal parts and merges adjacent sharps and flats into single notes.

Dr. Robert Smith begins at the other end, with the response of the ear. This is what makes his solution so interesting, and, indeed, more suggestive than the experimental organs of the nineteenth century, none of which has survived. The observation on which he based his system may be given in his own words, quoting from his Preface :—

“It is the transition from a better harmony to a worse, which chiefly gives the offence [to the ear]; as is evident to anyone that attends to a piece of music performed upon an instrument badly tuned. It follows then that the instrument would be better in tune, if all the consonances were made as equally harmonious as possible, though none of them were perfect.”

Elsewhere in the Preface he contrasts the fifth of mean-tone temperament, or of Huygen's system “resulting from the octave divided into 31 equal intervals,” with the major sixth. For mean-tone temperament the fifth is a quarter of a comma flat and the major sixth is a quarter of a comma sharp*. As a result the tempered major sixth would beat more than half as fast again as the tempered fifth. Dr. Smith tried the experiment on an organ, “and being solicitous to know the reason of that effect, that is, why the Vth and VIth major, when equally tempered † should differ so in their harmony, after various attempts I satisfied my curiosity.” This clearly indicates the relation he established between the beating of mistuned intervals and their dissonance. But proceeding on the assumption that tempered intervals would be equally harmonious if they produced beats at the same rate in adjacent concords, such as the fifth and major sixth, he devised what he described as the System of Equal Harmony for tuning keyboard instruments. He writes: “in the system of equal harmony the temperaments of the Vth, VIth and IIIrd are $-5/18$, $+3/18$, and $-2/18$ of a comma respectively and are proportional to the musical primes 5, 3, and 2.” Observe that the fifth is subjected to the most mistuning. This would horrify many nineteenth-century theorists, one of whose clichés was that equal temperament, which flattens the fifth by only one-eleventh of a comma but sharpens the major third and sixth by seven- and eight-elevenths of a comma respectively, “tampers most with those intervals which can best stand it.” Helmholtz brushed this claim aside by considering the dissonance of triads. Dr. Robert Smith, by theory and observation alike, demolished the assumption on which it rested. And, indeed, the fallacy of the cliché, like that of so many of the complacent assumptions of the nine-

* Phil. Mag. (ser. 7) xxxiv. p. 474 (1943).

† This accurate use of the words “equally tempered” suggests that “equal temperament” is an inappropriate term for a tuning which tempers different intervals very unequally.

teenth-century theorists, is obvious. For, if it were true, equal temperament must, from an early date, have superseded mean-tone temperament, which does not tamper at all with the major third, but flattens the fifth nearly three times as much as equal temperament does.

It would be impossible to summarize Dr. Smith's closely-reasoned argument, which he presents in a series of propositions informed from time to time by appeal to observation. Moreover, his premise about the dissonance of a mistuned concord is "speculative," and its interest is mainly historical. I can discern in his argument no physical or physiological reality, save in so far as his "Mathematical helps" * embody the periodicity of the partial tones of the notes forming the interval. That, however, is obviously why it led him to what we know to-day to be correct conclusions about the rate of beating of tempered intervals. More significant are the facts that it was his musical ear which impelled him to embark on his enquiries, and that he is insistent on the need to bring his conclusions to the test of practical observation, *e. g.* when he remarks (Proposition XVII.) "those only that have carefully attended to the beats of concords in tuning instruments" were "proper judges." The best that one can do by way of summary is to quote from his own Preface to the second edition :

"And lastly, as the harpsichord has neither strings nor keys for any of these sounds D \sharp , A \sharp , E \sharp , B \sharp , F $\sharp\sharp$, A \flat , D \flat , G \flat , &c., which yet are so often wanted that far the greater part of the best compositions cannot be performed without them, except by substituting for them E \flat , B \flat , F, C, G, G \sharp , C \sharp , F \sharp , &c., respectively, which, by differing from them by near a fifth part of the tone, make a very bad harmony ; and as the old expedient for introducing some of those sounds by inserting more keys in every octave, is quite laid aside by reason of the difficulty in playing upon them ; I have therefore invented a better expedient, by causing the several keys of those substitutes, E \flat , B \flat , F, C, G, G \sharp , C \sharp , F \sharp , &c., to strike either E \flat or D \sharp , B \flat or A \sharp , F or E \sharp , C or B \sharp , G or F $\sharp\sharp$, G \sharp or A \flat , C \sharp or D \flat , F \sharp or G \flat , &c. .

"For since both the sounds in any one of those couples are seldom or never used in any one piece of music, the musician by moving a few stops before he begins to play it, can immediately introduce that sound in each couple, which he foresees is either always or oftenest used in the piece before him.

"Two different constructions of those stops are here described, one of which is applicable at a small expense to any harpsichord ready made, and the other to a new harpsichord, and upon putting them both in practice, they have perfectly answered my expectation.

"Several properties and advantages of this changeable scale are described in the eighth Section. In a word, the very worst keys in the common defective scale [*i. e.* mean-tone tuning], by changing a few sounds are presently made as complete as the best in that scale, and more

* *Phil. Mag.* (ser. 7) xxxiv. p. 477 (1943).

harmonious too, because the changeable scale admits of the very best temperament, and, which is another advantage, will therefore stand longer in tune than the common scale which cannot admit that temperament.

"These improvements of the harpsichord, it is hoped, may encourage others to apply the like methods to the scale of the organ, which is equally capable of them and to greater advantages."

The harpsichord, as contrived by Hans Ruckers in the seventeenth century, had two manuals, to give the player control over the tone of the instrument (*e. g.* forte and piano). Dr. Smith's device involved the sacrifice of their advantages, and made use of a single manual whose digitals had each two jacks so arranged that, by the action of a stop, each digital could be made to pluck one or other of two strings which were originally proper to the two manuals respectively. The two series of strings were then tuned, by their rates of beating, in fifths, to give the two series of tones, differing by a diesis, which he sets out in the above quotation. The final paragraph is particularly interesting, as it anticipates a similar suggestion, by Helmholtz, for adapting the intonation of his experimental two-manual harmonium to the organ. The suggestion failed because of the cost of doubling the number of pipes, or alternatively because it would at least halve the number of speaking stops (Church authorities expected to secure as much variety of organ tone for their money as possible); and because the use of draw-stops to set the instrument for desired keys was inconvenient with tracker organs.

To complete the story it should be added that, having first determined his System of Equal Harmony, Dr. Robert Smith made computations which showed, in terms of the rate of beating, that his system was obtained, for all practical purposes, by dividing the octave into 50 equal parts, taking 8 of these for the whole tone and 5 for the limma (representing the diatonic semitone). This gave 3 divisions for the minor limma (the difference between the whole tone and the limma), and 2 for the diesis (the difference between the two limmas, *i. e.* between F# and Gb). In his Proposition XVII. there are two interesting tables, one comparing, by beats, the cycle of 50 with the system of equal harmony (it was for this comparison that he demanded skilled tuners as "proper judges"—see the quotation above); the other giving the temperaments of the Vth, VIth, and IIIrd in the cycles of 12 (equal temperament), 19 (see above), and 31 (Huygens). He makes no mention of Mercator's cycle, for, as we saw in my earlier paper, he accepted the identification of the major and minor tones, as a means of coping with the mutable notes of the musical scale, this being an essential and, as experience shows, satisfactory requirement of any practical temperament.

It is a curious, and it seems to me inexcusable, omission on Ellis's part (for he quotes from the book in another connexion) that, in his survey of temperaments in his Appendix to 'Sensations of Tone,' he should neglect to mention Robert Smith's cycle of 50. Perhaps the explanation is that Dr. Smith's device came into the field too late to be at all widely adopted. Within a short time of his death in 1768, the harpsichord

was superseded by the pianoforte. It was in 1783 that John Broadwood took out a patent for his reconstruction of Backer's original forte-piano. To-day Dr. Smith's work appears to call for more considered attention than the fruitless efforts of nineteenth-century theorists, none of whose organs proved sufficiently practicable to be copied; for, if provided with a reasonable range of speaking stops, not only would their cost have been prohibitive, but their complicated keyboards, applied to a 3-manual organ, would lie, I imagine, in a large measure beyond the reach of the organist's hands. Dr. Smith aimed at a practical device, and his solution had the merit that it sprang from a musical, not an arithmetical, conception. Yet I have so far come across only one brief mention, in any standard work, of his contribution to musical acoustics; in the article on beats in the first edition of Grove's 'Dictionary of Music,' in a statement that misses the point.

The real problem of the keyboard instrument remains to be fully solved. It is: Why does the average ear of the trained musician, which knows there is something wrong with the tuning of the harmonium, accept as satisfactory the tuning of the piano? Helmholtz offered a partial explanation. To many physicists the problem presents itself as being concerned with the tolerances of the ear. But, it seems to me, that it is not enough to explain why the ear fails to hear the mistuning involved in equal temperament on the piano. Something positive, not negative, is required; some examination of what the ear does hear, a problem in aural perception, belonging to "the psychological section of the theory of hearing" which Helmholtz left, in the main, as a study for his successors. Elsewhere* I have brought together some of the relevant musical evidence and put forward an analogy from visual perception which I owe to the kindness of Dr. Robert Thouless. These may help to define the problem still remaining. One further piece of evidence, which may be relevant, can be added. On the organ, there is a soft stop known as the *voix celeste* that is tuned slightly sharp on an associated stop with which it is used. The result is a vibrato-like sound, and the ear hears a definite note whose pitch lies between the pitches proper to the notes of the two sounding stops. Used for a reasonable limit of time the combined stops sound peculiarly sweet, and they appear to disguise in some measure the temperament used. This suggestion was confirmed, though not of set purpose, by a letter in 'The Musical Times' of January, 1941, from a practical organ-builder, who wrote: "A vibrato covers a multitude of sins, and the harshness of the harmonium tuning disappears when a well-tuned beating stop such as the Voix Celeste on a Mustel organ is used." The problem is: What intonation does the musical ear perceive in listening to such a stop or, alternatively, to the piano? Which is not quite the same thing as: Why does the ear tolerate equal temperament on the piano?

Gerrard' Cross,
Bucks.

* In the last essay in 'The Musical Ear' (Oxford University Press).

LXXV. *Preliminary Note on Condensation in the Form of Clouds and Dew.*

By J. W. ARCHBOLD, M.A., University College, London*.

[Received April 28, 1943.]

1. CLASSICAL investigations by Kelvin and Thomson on the condensation of water vapour on a single spherical water drop have shown that condensation takes place the more easily the larger the size of the drop, and that there is a minimum size of drop, corresponding to a given humidity, below which the drop will evaporate.

In considering the process of condensation by cloud formation we should therefore expect that large drops would grow at the expense of small ones. It is therefore important to assess the distribution of drops according to size, and so to exhibit the competition for growth amongst the drops. This paper represents a preliminary effort to do so.

The necessary approach, under quasi-static conditions at uniform temperature, has recently been given in general terms by J. Frenkel and W. Band (*J. Chem. Phys.* vii. 1939), and is applied here in more detail, taking account of the earth's gravitational field.

The positive conclusions are that

- (1) gravity appears to have no significant effect ;
- (2) in the absence of electrical charges, condensation tends to occur in large drops without any formation of cloud as the word is generally understood ; and the minimum size of drops to be expected in quantity is given by Kelvin's formula, which decides the minimum size of *single* drop on which condensation will occur ;
- (3) the distribution in size depends on the distribution of electrical charge : if all the drops carry the same numerical charge, there is still a tendency for condensation in large drops, whose minimum size is given by Thomson's formula which extends Kelvin's, but this is accompanied by a cloud of fine drops, whose sizes can be estimated by a formula given in the text, which may well be visible.

A negative conclusion is that ordinary cloud formation is probably associated with temperature variation and departure from equilibrium conditions ; though the relative effects of these two factors is a matter for further investigation. Perhaps a radically different technique is necessary to deal with these circumstances.

It may be remarked that, after deciding on a suitable model for an ice-crystal and obtaining an expression for its total energy, it should be possible to discuss on similar lines condensation in the form of a mixture of super-cooled water drops and ice-crystals, or of ice-crystals only.

In the last part of this note, it is shown that the sizes of dewdrops obey a similar distribution law to that for cloud drops, and that the

* Communicated by N. K. Johnson.

degree of affinity of water for the surface on which condensation takes place is of importance in regard to the passage beyond supersaturation to condensation.

2. We use classical statistical mechanics. The effect of the remaining drops on the potential energy of any one drop is assumed to be negligible. The cloud is, further, supposed to be in a state of macroscopic equilibrium. For simplicity, the drops are assumed to be spherical: we shall remark in place on the effects of non-sphericity.

We consider a fixed place, of volume V , in the atmosphere and the mean conditions inside. Each drop of radius r is to be regarded as an associated set of n molecules, where

$$nm = \frac{4}{3} \pi r^3 \rho, \quad \dots \dots \dots (1)$$

m being the mass of a molecule and ρ being the density of water. Let there be N_n such drops. In the case of partially formed drops which are too small to deserve the name "spherical," and in the case of large drops whose shape cannot be controlled effectively by surface tension (1) may be taken to define a vertical radius for a drop of n molecules.

The method of association is relevant only in so far as it affects the formula for the energy ϵ_n of the drop.

According to the law of mass action, we have

$$\frac{N_n}{(N_1)^n} = \frac{f_n}{(f_1)^n}, \quad \dots \dots \dots (2)$$

where f_n is the partition function for a drop consisting of n molecules.

3. We shall suppose that the whole volume V is at a uniform temperature T . Let x, y, z , be the co-ordinates of the centre of the drop, z being measured vertically upwards, and let p_x, p_y, p_z be the corresponding momenta. If we assume for the present that the rotational, vibrational and electrical contributions to the energy, and the energy of the medium resulting from the motion of the drop, can be neglected (the necessary corrections will be stated later), we have

$$f_n = \frac{1}{h^3} \int \exp \{-\epsilon_n/(kT)\} dx dy dz dp_x dp_y dp_z, \quad \dots \dots \dots (3)$$

where h is Planck's constant, k is Boltzmann's constant, the integration with respect to x, y, z is over the whole volume V , and with respect to each momentum is from $-\infty$ to $+\infty$; and

$$\epsilon_n = \frac{1}{2nm} (p_x^2 + p_y^2 + p_z^2) + \eta_n, \quad \dots \dots \dots (4)$$

where η_n is the potential energy of the drop.

Integrating with respect to the momenta, (3) becomes

$$f_n = \frac{1}{h^3} (2\pi nm kT)^{3/2} \int \exp \{-\eta_n/(kT)\} dx dy dz. \quad \dots \dots \dots (5)$$

Further discussion depends on evaluating η_n . For V we take an isothermal layer with base z_0 and top z_1 . Let σ be the surface tension

of water, λ be the latent heat of vaporization, and ρ_0 be the density of air, at temperature T . Then, if P is the atmospheric pressure at the base of the layer,

$$\eta_n = \frac{4}{3}\pi r^3 \{P - g\rho_0(z - z_0)\} + 4\pi r^2\sigma - \frac{4}{3}\pi r^3\rho\lambda + \frac{4}{3}\pi r^3g(\rho - \rho_0)z \quad (6)$$

for truly spherical drops, and

$$\eta_1 = mgz(1 - \rho_0/\rho) = \frac{4}{3}\pi r^3g(\rho - \rho_0)z/n. \quad (7)$$

For partially formed drops, the value of η_n is uncertain. At any rate, η_n must decrease steadily as n decreases and, apart from the gravitational term, be $O(n)$ for nearly-formed drops, while for drops in the very early stages of growth it must be $O(n^{2/3})$. Therefore, and by continuity, (6) should hold fairly well for nearly-formed drops, and for the smaller ones we should expect to replace the term $\frac{4}{3}\pi r^3\rho\lambda$ by one proportional to r^2 , representing the heat of adhesion to the nucleus of condensation.

The formula (6) will be inaccurate (increasingly so as r increases) for large drops because the surface energy is not then enough to keep the drops spherical. A reasonable estimate is that (6) holds well for $r < 0.1$, fairly well for $0.1 \leq r \leq 0.2$, and then less well. The general nature of the correction term comes from the fact that work must be done to make a large drop truly spherical; *i. e.*, we must add a positive term to the expression on the right of (6); *i. e.*, f_n , and consequently also N_n , will be less than given by (2), (5), (6). It is also evident, on elementary grounds, that N_n must decrease to zero as n increases, and in fact be zero before n is as large as the number of available molecules of water.

From (5), (6) and (7) we obtain

$$f_n = \frac{V}{h^3} (2\pi n m k T)^{3/2} \exp \left\{ - \left[\frac{4}{3}\pi r^3 (P + g\rho_0 z_0 - \lambda\rho) + 4\pi r^2\sigma \right] / [kT] \right\} \\ \times (z_1 - z_0)^{-1} \int_{z_0}^{z_1} \exp \left\{ - \frac{4}{3}\pi r^3 g(\rho - 2\rho_0)z / (kT) \right\} dz, \quad (8)$$

$$f_1 = \frac{V}{h^3} (2\pi m k T)^{3/2} (z_1 - z_0)^{-1} \int_{z_0}^{z_1} \exp \left\{ - \frac{4}{3}\pi r^3 g(\rho - \rho_0)z / (n k T) \right\} dz; \quad (9)$$

and therefore

$$\frac{N_n}{(N_1)^n} = \left\{ \frac{h^3}{V(2\pi m k T)^{3/2}} \right\}^{n-1} \cdot n^{3/2} \\ \times \exp \left\{ - \left[\frac{4}{3}\pi r^3 (P + g\rho_0 z_0 - \lambda\rho) + 4\pi r^2\sigma \right] / [kT] \right\} \cdot \phi_n, \quad (10)$$

where

$$\phi_n = \left\{ \frac{\theta(\rho - \rho_0)}{1 - e^{-\theta(\rho - \rho_0)}} \right\}^n \left\{ \frac{1 - e^{-n\theta(\rho - 2\rho_0)}}{n\theta(\rho - 2\rho_0)} \right\} \quad (11)$$

and

$$\theta = \frac{(z_1 - z_0)mg}{\rho k T} \quad (12)$$

Let $\tau = \theta(\rho - \rho_0)/(1 - e^{-\theta(\rho - \rho_0)})$; then, for small values of n , $\phi_n \approx \tau^n$; and for large values of n ($> 10^4$ effectively), $\phi_n \approx \tau^n / \{n\theta(\rho - 2\rho_0)\}$.

To fix ideas, let us take a layer 50 metres thick, and $T = 283$. We have $k = 1.3803 \times 10^{-16}$, $\rho = 0.99973$, $\rho_0 = 0.00125$, $m = 2.9966 \times 10^{-23}$. These give, approximately, $\theta = 0.003763$, $\theta(\rho - \rho_0) = 0.003758$, $\theta(\rho - 2\rho_0) = 0.003753$, $\tau = 1.0000$.

The tables available are unsatisfactory for these calculations. Actually τ is slightly greater than 1, so it is conceivable that τ^n may be quite large, for large values of n (up to 10^{21} , as we shall see, are involved). However, we shall proceed on the assumption that, effectively, $\tau = 1$; but we shall retain τ in the formulæ.

Normal variations in T make little difference to ϕ_n .

It will be convenient to continue the discussion on the lines of first ignoring gravity (or considering only small drops) and then seeing how results are affected by a gravitational field.

4. (a) Gravity absent.

For shortness, put

$$\alpha = \frac{V(2\pi mkT)^{3/2}}{h^3}, \quad \dots \dots \dots (13)$$

$$\beta = \frac{\sigma}{kT} \left(\frac{36\pi m^2}{\rho^2} \right)^{1/3}, \quad \dots \dots \dots (14)$$

$$\nu = \frac{N_1 h^3 \exp\{-m(P + g\rho_0 z_0 - \lambda\rho)/(\rho kT)\}}{V(2\pi mkT)^{3/2}}. \quad \dots \dots \dots (15)$$

Taking $T = 283$ again, $\sigma = 74$ dynes/cm. approximately. $P + g\rho_0 z_0$ differs little from the ground pressure, and we can take a sample value of 1010 millibars $= 1.01 \times 10^6$ dynes/sq. cm. But this is outweighed by the term $\lambda\rho$, which $= 591.4$ gr. cal. $= 2.5 \times 10^{10}$ ergs. Planck's constant $h = 6.54 \times 10^{-27}$ erg sec. With these values,

$$\beta = 8.8 \quad \text{and} \quad \nu = 2.98 \times 10^{18} N_1/V.$$

With these definitions we may write the distribution law in the form

$$N_n = \alpha \nu^n n^{3/2} e^{-\beta n^{2/3}}. \quad \dots \dots \dots (16)$$

It is not hard to see that the effect of introducing the rotational and vibrational energy of the drop, and the inertia of the medium, will be to give, instead of (16),

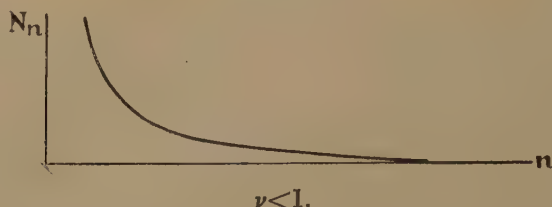
$$N_n = A \nu^n n^B e^{-\beta n^{2/3}}, \quad \dots \dots \dots (17)$$

where A and B ($> 3/2$) are independent of n , and ν , β are still given by (15), (14). It is immaterial which form we adopt for the conclusions which follow; we adopt (17).

For very small drops, the general effect will be to replace λ by 0 in (15) and to subtract a constant from β . In this connexion we note that, for these drops, $\nu = 1.4 \times 10^{-26} N_1/V$, pressure being now more significant than latent heat.

If $\nu < 1$, N_n decreases steadily and rapidly to zero as n increases. Such a distribution is stable.

Fig. 1.



If $\nu = 1$, N_n again decreases steadily and rapidly to zero after passing through a maximum when $n = 2\beta^{1/2}/3$ ($= 2$, approximately). This value is physically insignificant, and the distribution is essentially similar to the case $\nu < 1$; it is stable.

If $\nu > 1$, N_n decreases to a minimum, at $n = n_c$ say, and then steadily and rapidly increases (so long as the distribution law holds). The equation for n_c is

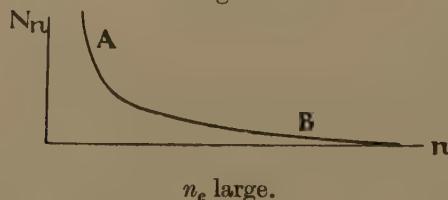
$$\log \nu = \frac{2\beta}{3n_c^{1/3}} - \frac{B}{n_c} \quad \dots \quad (18)$$

Since $n_c \gg 1$, a close approximation to (18) is

$$\log \nu = \frac{2\beta}{3n_c^{1/3}} \quad \dots \quad (19)$$

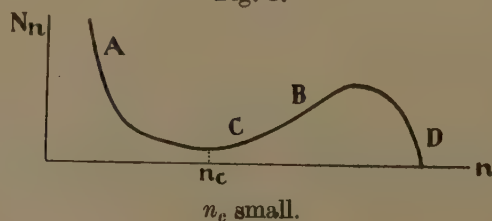
In this distribution, the number of drops of n molecules increases with n after n exceeds n_c , which suggests an unstable distribution. However, as already remarked, the total number of molecules is limited. The distribution must therefore be as follows if n_c is large (ν just greater than 1):—

Fig. 2.



and as follows if n_c is small:

Fig. 3.



In each case the stretch AB is supposed to represent the part to which the distribution law (17) applies. The case $\nu > 1$ thus falls into two essentially different types, one stable and the other meta-stable.

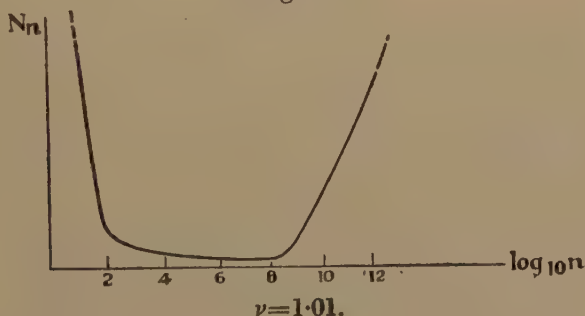
It is evident that $\nu < 1$ corresponds to an unsaturated vapour, which (as Frenkel remarks) consists partly of molecular clusters; $\nu = 1$ to a saturated vapour, ν just greater than 1 to a supersaturated vapour, and larger values of ν to cloud or drop condensation. The visible cloud essentially corresponds to the hump CD.

To get some idea of the actual shape of meta-stable curve, we may note that corresponding to the fact that drops remain fairly spherical with radii up to about 10^{-2} cm., the point B must have an n -co-ordinate of the order of 10^{17} . Consider the particular case $\nu = 1.01$, corresponding to $n_c = 2 \times 10^8$: we find rough values as follows, assuming (16) to hold:—

n	10^2	10^4	10^6	10^8	10^{10}	10^{12}	10^{15}	10^{18}
$\log_{10} \frac{N_n}{\alpha}$	-4.8	-1.7×10^3	-4×10^5	$+0.5 \times 10^6$	2.5×10^7	3.9×10^9	4.3×10^{12}	4.3×10^{15}

For $n > 10^{10}$, we have, practically, $N_n \propto \nu^n$; and obviously this will also apply to the law (17). There is a sharp fall in the number of drops just above molecular size until $n = n_c$, and then a sharp rise.

Fig. 4.



(b) Gravity acting.

Formulæ (16) and (17) become, for $n > 10^4$,

$$N_n = \alpha(\nu\tau)^n n^{1/2} e^{-\beta n^{3/2}} / [\theta(\rho - 2\rho_0)], \quad \dots \dots \dots (16')$$

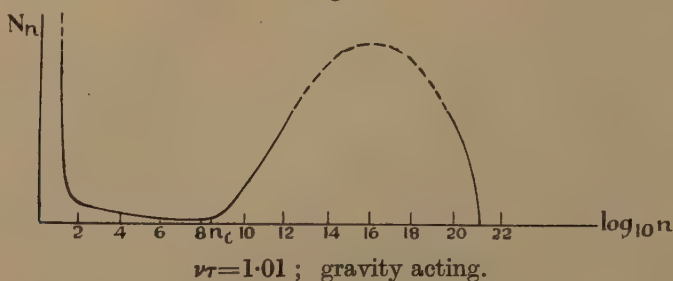
$$N_n = A(\nu\tau)^n n^{B-1} e^{-\beta n^{3/2}} / [\theta(\rho - 2\rho_0)]. \quad \dots \dots \dots (17')$$

Thus the effect of gravity is to replace the critical factor ν by $\nu\tau$; and therefore saturation now corresponds to $\nu\tau = 1$, or $\nu = 1/\tau$. However, as we have observed, $1/\tau$ is only very slightly less than 1.

To find where the graph turns down and reaches the n -axis we must have recourse to practical observations. These show that the maximum value of r is about 2.5×10^{-1} cm., which corresponds to $n = 2 \times 10^{21}$.

Thus, with a suitable scale for N_n , the graph of N_n in the meta-stable conditions is roughly

Fig. 5.



To complete the picture, we have, approximately,

$$n_c = 2 \times 10^{3x+2} \quad \text{if } \nu\tau = 1 + 10^{-x}, \quad \text{for } x \geq 2.$$

Thus there will be no hump if $1 < \nu < 1 + 5 \times 10^{-7}$, roughly. This range is, however, too narrow because, even if ν is slightly greater than $1 + 5 \times 10^{-7}$, the hump will be too small for notice.

5. The conclusion to be drawn is that, under equilibrium conditions, or under conditions of very slow change, condensation begins in the form of *large* drops; and that, for any given value of $\nu > 1$, the minimum size of drop is given by

$$\log(\nu\tau) = \frac{2\beta}{3n_c^{1/3}}. \quad \dots \dots \dots (19')$$

In terms of the critical radius r_c , (19') is the same as

$$\log(\nu\tau) = \frac{2m\sigma}{\rho k T r_c}. \quad \dots \dots \dots (20)$$

The number $\nu\tau$ has a simple physical significance. Let P_1 denote the vapour pressure of the single free molecules; then

$$P_1 V = N_1 k T, \quad \dots \dots \dots (21)$$

and therefore

$$P_1 = \frac{kT(2\pi mkT)^{3/2}}{h^3} \cdot \exp\{m(P + g\rho_0 z_0 - \lambda\rho)/(\rho kT)\}. \quad \dots \dots (22)$$

Let P_s be the value of P_1 when $\nu\tau = 1$; P_s may be called the monomolecular saturation vapour pressure. Then

$$P_s = \frac{kT(2\pi mkT)^{3/2}}{h^3} \cdot \exp\{m(P + g\rho_0 z_0 - \lambda\rho)/(\rho kT)\}, \quad \dots \dots (23)$$

and

$$\nu\tau = P_1/P_s. \quad \dots \dots \dots (24)$$

We can therefore write (20) in the form

$$\log \frac{P_1}{P_s} = \frac{2m\sigma}{\rho k T r_c}. \quad \dots \dots \dots (25)$$

Assuming spherical drops only to be possible, neglecting gravity, and considering the limit as $\nu \rightarrow 1$, i. e. as $n_c \rightarrow \infty$, we see that a saturated

water vapour will remain in equilibrium with a plane water surface. Thus (25) amounts to Kelvin's well-known formula.

If $P_s = P'_s$ when $T = T'$, we have, from (23),

$$\log \frac{P_s}{P'_s} = \frac{5}{2} \log \frac{T}{T'} + \frac{m(P + g\rho_0 z_0 - \lambda\rho)}{\rho k} \left(\frac{1}{T} - \frac{1}{T'} \right) \quad (26)$$

(treating ρ as constant in the range T to T'). Over ordinary temperature ranges $\log (T/T') \approx 0$; also $P + g\rho_0 z_0$ is small compared with $\lambda\rho$; hence, approximately,

$$\log \frac{P_s}{P'_s} = -\frac{m\lambda}{k} \left(\frac{1}{T} - \frac{1}{T'} \right) \quad (27)$$

This is a well-known equation which can be derived from the equation of state and the Clausius-Clapeyron equation when only a single drop is present among the free molecules.

6. We now consider electrical effects. Without needing to go into the question of how a drop is to acquire an electrical charge, we can consider the effects of two simple alternative hypotheses about the distribution of charge.

First, let us suppose that each drop carries the same charge q . This amounts to adding a term $q^2/(2r)$ to the expression for the potential energy. The distribution law becomes

$$N_n = A v^n n^B \exp \left\{ -\beta n^{2/3} - \frac{q^2}{2kT} \left(\frac{4\pi\rho}{3nm} \right)^{1/3} \right\} \cdot \phi_n \quad (28)$$

The term $\frac{q^2}{2kT} \left(\frac{4\pi\rho}{3m} \right)^{1/3}$ has the value $7.8 \times 10^{20-2x}$, where $q = 10^{-x}$; for shortness, denote this term by C .

The criterion for saturation is unaltered. The stationary values of N_n/ϕ_n are given by

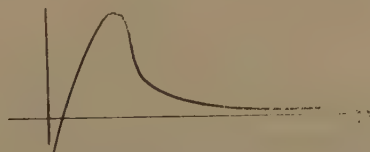
$$\log (v) = \frac{2\beta}{n^{1/3}} - \frac{B}{n} - \frac{C}{3n^{4/3}} \quad (29)$$

For very small charges C is small and, practically, there is only one stationary value given by (19), corresponding to a minimum value of N . For larger values of q , C is large, and (bearing in mind that we are limited to $n < \text{about } 2 \times 10^{21}$), (29) becomes effectively

$$\log (v) = \frac{2\beta}{n^{1/3}} - \frac{C}{3n^{4/3}} \quad (30)$$

The graph of the right-hand side of (29) is roughly :

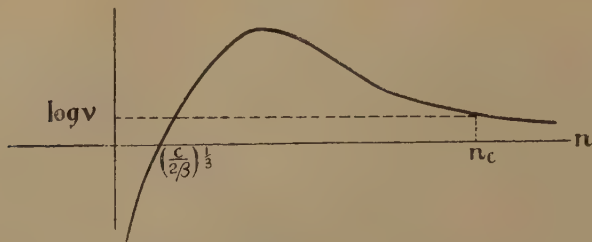
Fig. 6.



Graph of $\frac{2\beta}{n^{1/3}} - \frac{B}{n} - \frac{C}{3n^{4/3}}$, C small

and of the right-hand side of (30) is :

Fig. 7.



Graph of $\frac{2\beta}{n^{1/3}} - \frac{c}{3n^{4/3}}$, C large.

In all cases, the graph crosses the n -axis steeply ; hence the roots of (30) are $n=n_c$, corresponding to a minimum value of N_n , and, approximately, $n=\left(\frac{c}{2\beta}\right)^{1/3}$, corresponding to a maximum value of n .

The minimum size of large drop is therefore given by Thomson's formula

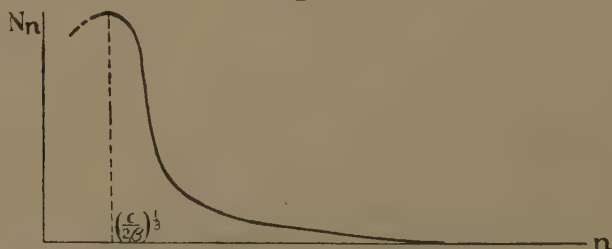
$$\log \frac{P_1}{P_s} = \frac{m}{\rho k T} \left(\frac{2\sigma}{r_c} - \frac{q^2}{8r_c^4} \right), \quad \dots \dots \dots (31)$$

and does not differ greatly from that given by Kelvin's formula. However, the effect of the charge is also to produce a concentration of small drops around $n=\left(\frac{c}{2\beta}\right)^{1/3}$; and the larger the charge, the more visible do these drops become. For example, with $q=0.001$ e.s.u., we have $n=3.5 \times 10^4$, or $r=6.3 \times 10^{-7}$; and with $q=0.01$ e.s.u., we have $n=1.6 \times 10^5$, or $r=1.05 \times 10^{-6}$.

Thomson's formula gives a larger value than Kelvin's for r_c . Thus we may say that, when a charge is present, less condensation takes place in the form of large drops than when no charge is present. The difference is represented by a cloud of small drops.

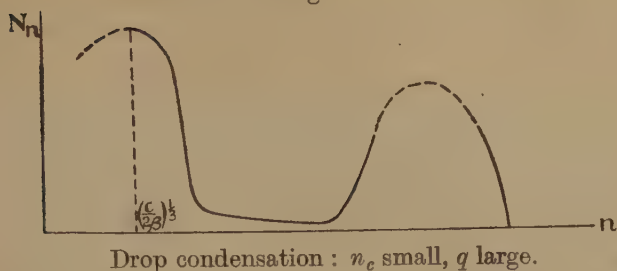
The distribution curves are

Fig. 8.



Supersaturation : n_c, q both large.

Fig. 9.



A special case arises if the value of q is such as to make the two roots of (30) equal. The drop-condensation curve then has an inflexion at $n=n_c$. When this occurs,

$$q^2 = \frac{128\sigma^4}{81} \left(\frac{m}{\rho k T \log v} \right)^3.$$

For larger values of q , both roots are unreal. N_n then increases steadily until the top of the hump is reached, and supersaturation is impossible. However, with the figures given, we should require $q > 0.054$ e.s.u., which is too large for ordinary atmospheric processes.

7. A reasonable alternative hypothesis is to suppose that the drops carry a charge q per unit area. This amounts to adding a term $(4\pi^2 q)^2 / (2r)$ or $\frac{4}{3}\pi r^3 \cdot 6\pi q^2$ to the expression for the potential energy. This is equivalent to increasing the atmospheric pressure by $6\pi q^2$. The effect of this is negligible.

Since, in both hypotheses, only q^2 is involved, the sign of the allotted charges can be arbitrary. But some doubt arises as to the effect on potential energy of the interaction between the drops, an effect which the theory assumes to be nil.

8. The Sizes of Dewdrops.

Considerations, similar to those above, apply to the distribution in size amongst dewdrops, probably with more force since equilibrium conditions may more nearly be realized.

For a simple theory, we consider a horizontal plane surface and suppose that the drops in contact with it are sections of spheres.

Fig. 10.



Let σ be the boundary tension between water and air, and σ' be that between water and surface; then $\sigma' = \sigma \cos \psi$.

On the basis of the preceding theory we shall neglect gravity. And

if we suppose that the boundary surface is of conducting material and earthed, we can ignore electrical effects. Then it is easy to show that

$$\eta_n = \left\{ \frac{4}{3} \pi r^3 (P - \lambda \rho) + 4 \pi r^2 \sigma \right\} \omega, \quad (32)$$

where
$$\omega = \frac{1}{4} (2 + 3 \cos \psi - \cos^3 \psi); \quad (33)$$

and
$$nm = \frac{4}{3} \pi r^3 \rho \omega. \quad (34)$$

Therefore the distribution law becomes

$$N_n = \alpha \nu^n n^{3/2} e^{-\beta_0 n^{1/3}}, \quad (35)$$

where α, ν are given by (13), (15) respectively, and

$$\beta_0 = \beta \omega^{1/3}. \quad (36)$$

The criteria for the different degrees of stability are unchanged. In the meta-stable state, the minimum size of large drop is given by

$$\log \nu = \frac{2\beta_0}{3n_c^{1/3}}, \quad (37)$$

which may still be expressed by Kelvin's formula

$$\log \nu = \frac{2m\sigma}{\rho k T r_c}.$$

As ψ increases from 0 to π , ω decreases from 1 to 0. Thus the closer the affinity of the water for the surface, the smaller will be the actual volume of the minimum large drops. Practically, this means that on a surface with high affinity for water, the phenomenon of super-saturation will be almost unnoticeable: thus condensation will start on such a surface before it would do so in free air.

9. Whether we are concerned with condensation in the free air or on a surface, n_c is proportional to $\sigma^{1/3}$. Now the surface tension of water is reduced, in most cases, when a substance is dissolved in it; the consequent reduction in n_c , therefore, favours the development of a hump in the distribution curve: that is, the super-saturation condition is likely to be passed through, during a process of cooling, more quickly than with pure water-drops. Therefore, the chemical nature of the nucleus of condensation may be expected to operate in the early stages of the growth of the drop if the nucleus is soluble, in addition to any purely adhesive power it may have in regard to water molecules.

In a polluted atmosphere, the soluble particles and gases must all help the process of condensation. It is to be noted that though a particular substance may be ineffective as a condensation nucleus, yet it may become effective in the growth of the drop, once the drop has been formed. It is therefore to be expected that, other factors being equal, there will be more large drops in a town fog than in a country fog. And a similar remark applies to dew formation on a dirty surface; though here a reduction in σ , due to solubility, might well be counterbalanced by a diminution of ψ .

Lagos,

January 10, 1943.

LXXVI. *Viscosity at the Boiling Point.—The Rheochor.*

By J. NEWTON FRIEND and WILLIAM D. HARGREAVES*.

[Received June 18, 1943.]

CHEMISTS have for long sought, but with indifferent success, some expression that would correlate the composition and structure of a liquid with its viscosity. In 1909, Dunstan and Thole⁽¹⁾ showed that, in certain homologous series, a linear relation holds between the molar weight, M , and $\log \eta$ at constant temperature, where η is the viscosity. They evaluated several atomic, group and structural constants; but the value of these is severely restricted because any selected temperature can only lie within the range of liquidity of a relatively small number of substances. Linke⁽²⁾, using Sheppard's equation⁽³⁾, $\log \eta = A + \frac{B}{T}$, showed that the constants A and B vary linearly with the number of carbon atoms in homologous series; but this again is too restricted to be of general use.

What is really needed, therefore, is some function of viscosity that is practically independent of the temperature over a considerable range. In a series of papers, Herz⁽⁴⁾ has shown that $M\eta^{1/2}/D$ is constant for non-associated liquids over such a range, D being the liquid density.

That a connexion should exist between viscosity and surface tension, σ , has long been realized. Sharma⁽⁵⁾ pointed out that $\log \sigma$ is, in certain cases, a linear function of $\log \eta$ over a restricted range of temperature. A neat expression was obtained by Silverman and Roseveare⁽⁶⁾, who combined Macleod's equation⁽⁷⁾ with the earlier one of Batschinski⁽⁸⁾ and obtained

$$\sigma^{-1/2} = A/\eta + B,$$

where A and B are constants.

Attention has already been directed⁽⁹⁾ to the fact that, over a considerable range of temperature from just below the boiling point downwards, $\sigma/\eta^{1/2}$ is constant for non-associated substances. Thus, expressing viscosities in millipoises for ease of calculation, we obtain:

	$k = \sigma/(10^3 \eta)^{1/2}$				
Temp. °C.	10	20	40	60	80
Benzene	11.0	11.4	11.9	12.0	11.8
Methyl butyrate.....	10.2	10.4	10.7	10.7	10.5
Bromine	13.0	13.2	13.3	13.3	—

* Communicated by the Authors.

The value of k varies considerably with the nature of the substance so that k is by no means a universal constant. Hence for a given liquid k will remain constant only over a temperature range in which the structure remains substantially the same.

In 1924, Sugden⁽¹⁰⁾ showed that $M\sigma^{\frac{1}{2}}/D$, which he named the parachor, is an additive quantity independent of the temperature. Replacing $\sigma^{\frac{1}{2}}$ by $\eta^{\frac{1}{2}}$ gives $M\eta^{\frac{1}{2}}/D$, termed the rheochor⁽¹¹⁾, which must likewise be independent of the temperature over the range for which k is constant.

With all liquids, whether regarded as associated normally or not, R rises as the melting point is approached. This is probably due to the grouping or orientation of the molecules which occurs prior to solidification. On the other hand, R again rises with the temperature from the boiling point upwards. At these temperatures the vapour density, d , becomes appreciable, and replacement of D by $D+2d$ gives a constant value for R over a greater temperature range. This is purely an empirical observation. Thus :

$$R = M(10^3\eta)^{\frac{1}{2}}/(D+2d).$$

Temp. °C.	60	80	100	140	180	240
Benzene (B. pt. 80°)	110.4	110.4	109.8	109.8	109.2	—
Chlorobenzene (B. pt. 131.7°) ..	129.8	129.4	129.5	130.1	130.6	131.5
Carbon tetrachloride (B. pt. 76.75°)	126.0	125.3	125.1	125.2	124.2	—
Ethyl alcohol (B. pt. 78.3°) ..	76.1	75.0	73.9	72.8	—	—

For associated substances, *e. g.* alcohol, R falls with rise of temperature. For such substances the boiling point is a convenient temperature for a comparison of R . All subsequent values for R in this paper, unless otherwise stated, refer to the boiling point.

For monomeric R is a definite fraction of the molar critical volume V_c . Thus :

	R .	V_c .	R/V_c .
Chlorine	27.7	61.9	0.448
Sulphur dioxide	52.4	122.1	0.429
Heptane	175.8	415.5	0.423
Benzene	110.4	257.0	0.429
Chlorobenzene	130.0	307.9	0.422
Chloroform	99.5	231.4	0.430
Ethyl acetate	118.5	286.0	0.414
		Mean....	0.426

R is additive, and analogous isomerides yield similar values. Thus :

$\text{CH}_2\text{Cl}.\text{CH}_2\text{Cl}$	101.5	$\text{C}_6\text{H}_5.\text{C}_2\text{H}_5$	153.6
$\text{CH}_3.\text{CHCl}_2$	101.3	$\text{C}_6\text{H}_4(\text{CH}_3)_2$	
$\text{C}_2\text{H}_5.\text{O}.\text{C}_2\text{H}_5$	115.5	ortho	153.8
$n\text{C}_3\text{H}_7.\text{O}.\text{CH}_3$	115.1	meta	153.6
		para	154.0

Isomerides of different types may, however, yield somewhat different values. Thus :

$\text{C}_6\text{H}_5\text{COOH}$	99.1	$\text{C}_6\text{H}_5\text{NH}\cdot\text{CH}_3$	144.0 ¹²
HCOOC_2H_5	95.9	$p\text{CH}_3\cdot\text{C}_6\text{H}_4\cdot\text{NH}_2$	147.3 ¹²

For monomeric substances, increase of external pressure up to about 6000 kilos/cm.² at constant temperature appears to have only a slight effect upon R. Thus, using Bridgman's⁽¹³⁾ data at 30° C :

Pressure, kilo/cm. ²	1	500	2000	4000	6000
Acetone	85.2	84.3	83.2	83.2	84.3
Carbon di-sulphide	71.1	70.1	69.1	68.8	69.2
Ethyl bromide	90.3	88.7	88.0	88.6	90.5

There is a slight initial fall to a minimum at about 4000 kilo/cm.², followed by a rise which becomes rapid at still higher pressures, when deformation of the molecules may occur.

Atomic, Group and Structural Rheochors at the Boiling Point.

For the evaluation of R the viscosities of the selected liquids at the boiling point, when not given by the authors concerned, have been calculated with the aid of Sheppard's equation. It frequently happens, however, that for many particularly suitable liquids data are not available sufficiently near the boiling point to permit of accurate extrapolation. Further, viscosity data are not always accompanied by densities. Since densities recorded by different investigators often show considerable variation, it is difficult to decide which data to select. The experimental work detailed below was therefore undertaken to fill in certain gaps, and the results have been used, together with other available data, in determining the following values for R :—

$$R = M(10^3\eta)^{1/2} / (D + 2d).$$

CH_2	23.8	H (in ordinary C—H)	5.5
C	12.8	H (in C.OH)	10.0
O (Etheric)	10.0	Cl	27.3
CO_2O (acids and esters)....	36.0	Br	35.8
		H in HCl	9.7
		H in HBr	12.6
		H in HI	15.0
C_6H_5 attached to alkyl radicle			100.7

The value for CH_2 is the mean obtained from some 60 liquids, mostly members of various homologous series, adjacent members of which differ only in their CH_2 content. The value for hydrogen varies greatly with the element or group to which it is attached.

We should expect this, for when the attractive force between two atoms varies, so will the distance between their nuclei, and hence the apparent volume and rheochor of the added element. This is well illustrated by the hydrogen halides. The attraction for hydrogen

becomes progressively weaker as we pass from Cl to I, and the value for R (hydrogen) rises. Thus :

	R (for HX).	R (for H).	<i>f</i> .	<i>f</i> R(for H).
HCl	37.0	9.7	1.59	15.4
HBr	48.4	12.6	1.30	16.4
HI	62.6	15.0	1.00	15.0

The relative "restoring force," *f*, between the hydrogen and halogen nuclei, is seen (last column) to be inversely proportional to R (for H). *f* has been calculated, using Czerny's observation that it is inversely proportional to the square of the mean distance between the atomic nuclei.

The evaluation of chlorine is typical and illustrates the type of agreement found.

Compound.	R.	C_nH_m .	Cl.
CH_2Cl_2	76.2	23.8	2×26.2
$CHCl_3$	99.5	18.3	3×27.1
CCl_4	125.4	12.8	4×28.2
$CH_2Cl.CH_2Cl$	101.5	47.6	2×27.0
nC_3H_7Cl	104.2	76.9	27.3
Mean			27.3

Interesting confirmation is afforded by subtracting half di-allyl, $CH_2:CH.CH_2.CH_2.CH:CH_2$, from allyl chloride, $CH_2:CH.CH_2Cl$, namely $95.2 - \frac{1}{2} 135.8$ or 27.3. This gives the same value for chlorine and is entirely independent of the values assigned to C and H. For free Cl, $R=27.7$.

CCl_4 gives a high value, probably because so many Cl atoms are attached to the same C. As shown later, $CBBr_4$ also yields a correspondingly high value.

For water R falls steadily as the temperature rises to the boiling point, when it becomes substantially constant at approximately the value given by the sum of $2H+O$. Thus :

Temp. °C.	20	60	100	120	140	153
R	24.1	22.2	21.4	21.1	21.1	21.0

Experimental.

Viscosities were determined with standard viscometers of the Ostwald type, Hosking's absolute values for water being used in the calculations. No correction has been applied for the kinetic energy of efflux as the time of outflow was never less than 360 seconds. Experiments at the higher temperatures were conducted in a 4-litre bath of sulphuric acid, the thermo-regulator being filled with mercury and controlling the temperature to within $0.1^\circ C$.

The three alkyl bromides were prepared from the re-distilled alcohols by Kamm and Marvel's method ⁽¹⁴⁾; they were kept over calcium chloride and finally distilled direct into the viscometer and density bottle.

The remaining substances were obtained direct from British Drug Houses. The carbon tetrabromide was kept in a desiccator for some days and used direct after melting and filtering through a sintered glass funnel. It was not distilled. The other substances were dried and distilled direct into the viscometer and density bottle.

In all cases density values refer to water at 4° C. The values of A and B refer to Sheppard's equation and to the linear part of the curve obtained on plotting $\log 10^3\eta$ against $1/T$. The boiling point values for η and D, obtained by extrapolation, are given in parentheses.

n-Butyl bromide.—Only the *iso* derivative was studied by Thorpe and Rodger⁽¹⁵⁾. B. pt. 101.5 to 102.0° C. at 751.3 mm. $A = -0.6437$; $B = 422.5$.

Temp.	15.8	49.6	69.5	82.4	89.1	95.0	102
D	1.2752	1.2304	1.2027	1.1823	1.1724	1.1631	(1.1530)
1000 η	6.597	4.637	3.909	3.513	3.330	3.200	(3.04)

$R = 135.5$; calculated 136.5.

n-Amyl bromide.—B. pt. 127.0° to 128.0° C. at 749.9 mm. $A = -0.7190$; $B = 479.6$, from 56° upwards.

Temp. ..	14.4	56.0	71.3	86.7	103.9	119.0	128.0
D	1.2247	1.1722	1.1527	1.1321	1.1084	1.0875	(1.0752)
1000 η ...	8.913	5.546	4.785	4.155	3.595	3.194	(3.00)

$R = 159.8$; calculated 160.3.

iso-Amyl bromide.—B. pt. 120.0° to 120.5° C. at 755.2 mm. $A = -0.8228$; $B = 508.9$, from 50° upwards.

Temp. ...	12.5	50.0	71.6	84.7	96.2	111.2	120.5
D	1.2172	1.1702	1.1395	1.1221	1.1054	1.0845	(1.0718)
1000 η ...	8.596	5.662	4.475	3.992	3.600	3.173	(2.956)

$R = 160.0$; calculated 160.3.

Methylene di-bromide.—B. pt. 97.2° C. at 754.2 mm. $A = -0.3420$; $B = 411.8$ from 45° upwards.

Temp.	18.6	45.0	55.5	65.1	74.7	85.0	91.4	97.2
D	2.4922	2.4259	2.4006	2.3758	2.3504	2.3243	2.3055	(2.2905)
1000 η	10.54	8.087	7.359	6.792	6.286	5.787	5.534	(5.32)

$R = 93.1$; calculated 95.4.

Dunstan⁽¹⁶⁾ and co-workers gave $1000\eta = 12.25$ at 25° C., with $D = 2.4080$; our values obtained by interpolation are $1000\eta = 9.863$, $D = 2.4771$. The discrepancy is great, but our density value is practically identical with that given by Perkin⁽¹⁷⁾, viz. 2.47745 at 25°.

Bromoform.—A few scattered viscosity data below 80° C. are given in

the literature⁽¹⁸⁾. B. pt. 147.7° at 743.9 mm. $A = -0.3692$; $B = 490.9$, from 60° upwards.

Temp. .	16.3	62.8	90.3	103.5	123.4	141.0	147.7
D	2.8927	2.7735	2.7022	2.6654	2.6113	2.5569	(2.5500)
1000 η .	21.84	12.40	9.663	8.618	7.367	6.577	(6.276)

$R = 124.0$; calculated 125.7.

Carbon tetra-bromide.—The white crystals melted to a pale yellow liquid; the colour deepened as the temperature rose. At 170° C. bubbles formed in the viscometer due to decomposition. B. pt. taken as 190° (from literature). $A = -0.9593$; $B = 877.5$.

Temp.	100.7	124.9	139.2	147.5	155.2	163.1	190
D	2.9533	2.9004	2.8609	2.8355	2.8094	2.7831	(2.6970)
1000 η	24.44	17.725	14.81	13.40	12.39	11.285	(8.63)

$R = 160.0$; calculated 156.0.

The discrepancy is high; this may be due to so many Br atoms being attached to the same C atom, since CCl_4 also gives a high result; further, the difference between CBr_4 and CCl_4 is $160.0 - 125.4 = 4 \times 8.65$, which agrees closely with the observed mean difference between Br and Cl, viz. 8.5.

Methyl ethyl ketone was studied by Thorpe and Rodger. It boiled between 79° and 83° C. and was too small in quantity to fractionate further. Our fraction boiled at 79.4° to 79.8° C. at 747.5 mm. $A = -0.7443$; $B = 399.3$.

Temp.	25.1	40.0	54.9	69.9	79.8
D	0.7994	0.7832	0.7673	0.7511	(0.7400)
1000 η	3.938	3.398	2.979	2.630	(2.444)

$R = 108.2$; calculated 108.4.

Our viscosity-temperature curve crosses that of Thorpe and Rodger at 45° ; above this temperature the results are closely similar.

Propionic acid.—Thorpe and Rodger did not determine the densities of their own specimen. Dunstan⁽¹⁹⁾ stated that he obtained similar results for the viscosity but again gave no densities. Our fraction boiled at 137.5° to 138.5° C. at 757.1 mm. $A = -0.8570$; $B = 568.2$ from 90° upwards.

Temp. .	15.8	78.9	95.9	108.5	119.2	133.2	138.5
D	0.9985	0.9301	0.9126	0.8990	0.8869	0.8706	(0.8650)
1000 η .	12.28	5.645	4.838	4.305	3.910	3.485	(3.345)

Our results are approx. 0.1 unit higher than those of Thorpe near the B. pt. $R = 99.1$; calculated 99.1.

n-Butylformate.—B. pt. 105.5° C. at 743.5 mm. $A = -0.9100$;
 $B = 512.2$.

Temp. ...	16.3	52.13	77.9	86.9	97.15	105.5
D	0.8946	0.8569	0.8298	0.8196	0.8078	(0.7982)
1000 η ...	7.244	4.585	3.545	3.259	2.977	(2.77)

$R = 144.0$; calculated 143.7.

Gartenmeister's⁽²⁰⁾ result at 20° for the viscosity, viz. 5.627, is low, that calculated from the above data being 6.89.

iso-Butyl formate.—B. pt. 95.5° to 97.0° C. at 739.9 mm. $A = -0.8920$;
 $B = 500.0$.

Temp. ...	16.05	52.3	63.4	74.6	88.0	97.0
D	0.8861	0.8474	0.8354	0.8239	0.8085	(0.7982)
1000 η ...	6.877	4.388	3.922	3.517	3.108	(2.88)

$R = 144.8$; calculated 143.7.

The viscosity calculated at 20°, viz. 6.52, lies between that of Gartenmeister, 6.65, and of Matthews and Faville⁽²¹⁾, 6.44. Iso-derivatives usually give a slightly higher value for R than the normal.

n-Amyl acetate.—Several data have been published⁽²²⁾ for "amyl acetate," but the various authors have not specified which acetate they used, nor the particular alcohol from which it was prepared. As the ester was used by them as a solvent, relative rather than absolute values were important.

B. pt. 144.0° to 144.5° at 737.2 mm. $A = -0.9620$; $B = 566.3$.

Temp..	15.3	42.4	63.0	81.7	100.45	119.4	135.0	144.5
D	0.8805	0.8544	0.8350	0.8165	0.7975	0.7783	0.7619	(0.7520)
1000 η .	10.070	6.741	5.271	4.289	3.594	3.031	2.668	(2.48)

$R = 192.0$; calculated 189.8.

iso-Amyl acetate, $\text{CH}_3\text{COOCH}_2\text{CH}_2\text{CH} \begin{smallmatrix} \text{CH}_3 \\ \text{CH}_3 \end{smallmatrix}$ —B. pt. 140.0° C. at
 760 mm. $A = -0.9578$; $B = 559.5$.

Temp. ...	12.5	59.3	100.9	120.4	135.0	140.0
D	0.8808	0.8349	0.7925	0.7708	0.7560	(0.7510)
1000 η ...	10.13	5.325	3.463	2.904	2.588	(2.49)

$R = 192.2$; calculated 189.8.

References.

- (1) Dunstan and Thole, *J. Chim. Phys.* vii. p. 210 (1909) ; *J. Chem. Soc.* ciii. p. 127 (1913).
- (2) Linke, *Z. physikal Chem.* A, clxxvii. p. 227 (1940). See Copley, 'Nature,' exlvii. p. 207 (1941).
- (3) Sheppard, *J. Rheology*, p. 208 (1930) ; Sheppard and Houck, *ibid.* p. 349 ; Andrade, 'Nature,' cxv. pp. 309, 582 (1930).

- (4) Herz, *Z. Elektrochem.* xxxvi. p. 454 (1930).
- (5) Sharma, *J. Indian Chem. Soc.* ii. p. 310 (1925).
- (6) Silverman and Roseveare, *J. Amer. Chem. Soc.* liv. p. 4460 (1932).
- (7) Macleod, *Trans. Farad. Soc.* xix. p. 38 (1923).
- (8) Batschinski, *Z. physikal. Chem.* lxxxiv. p. 643 (1913).
- (9) Friend, 'Nature,' cl. p. 432 (1942).
- (10) Sugden, *J. Chem. Soc.* cxxv. p. 1185 (1924).
- (11) Friend, *loc. cit.*
- (12) From unpublished data.
- (13) From Int. Critical Tables, vols. iii. and vii.
- (14) Kamm and Marvel, *J. Amer. Chem. Soc.* xlii. p. 299 (1920).
- (15) Thorpe and Rodger, *Phil. Trans.* clxxxv. p. 476 (1895).
- (16) Dunstan, Hilditch and Thole, *J. Chem. Soc.* ciii. p. 133 (1913).
- (17) Perkin, *ibid.* xlv. p. 520 (1884).
- (18) Beck, *Z. physikal. Chem.* lviii. p. 425 (1907); Drucker and Kassel, *ibid.* lxxvi. p. 367 (1911); Timmermans and Martin, *J. Chim. Phys.* xxv. p. 411 (1928); Trew, *Trans. Faraday Soc.* xxviii. p. 509 (1932).
- (19) Dunstan, *J. Chem. Soc.* cvii. p. 667 (1915).
- (20) Gartenmeister, *Zeitsch. physikal. Chem.* vi. p. 524 (1890).
- (21) Matthews and Faville, *J. Phys. Chem.* xxii. p. 1 (1918).
- (22) Baker, *J. Chem. Soc.* ciii. p. 1662 (1913); Thole, *ibid.* xcvi. p. 2601 (1910).

The Technical College,
Birmingham.

LXXVII. Notices respecting New Books.

Fundamental Principles of Electric and Magnetic Circuits. By Prof. F. A. FISH. [Pp. xiii+229.] (London: McGraw-Hill Publishing Co., Ltd., 1940. Price 20s.)

THIS useful textbook should be of great help to the undergraduate student of electrical engineering. It begins at the very beginning with a chapter devoted to the explanation of the fundamental concepts of mechanics—acceleration, mass, force, work, energy and power; Chapter 2 deals similarly with the fundamentals of electricity and magnetism; Chapter 3 with electric circuits; Chapter 4 with electromagnetism; Chapter 5 with electrostatics. This development takes up a little more than half the book, and, the foundations being thus well and truly laid, the rest of the book is given up to an elementary and very clear discussion of sine wave alternating currents, non-harmonic waves, and polyphase currents.

The book provides an excellent first-year course, and the student who uses it will find he has nothing to unlearn when he comes to study the higher branches of power machinery and transmission.

The printing and general production are, as usual, first-rate.

[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]

LXXVIII. *Difficulties in Dirac's Representation Theory.*

By H. A. C. DOBBS*.

[Received March 11, 1943.]

PART I.

I SHALL be chiefly concerned in this note to consider a paradox which is, I think, to be found in Dirac's formulation of the principles of quantum theory as set out in his book 'Quantum Mechanics' (Oxford University Press, second edition). It has not, as far as I know, been fully discussed before, and Dirac himself does not seem to have been aware of its existence, though it appears to be a serious blemish in his general representation theory.

In 'Quantum Mechanics' Dirac is concerned to provide an adequate account of the nature of the fundamental laws and conditions obtaining in atomic systems. This object has been the common aim of all modern quantum theories, since experimental observation first revealed the utter inadequacy of the principles and concepts of classical mechanics to explain the distribution of the frequencies of the electromagnetic radiation resulting from the oscillatory disturbances in an atomic system.

For example, the classical physical theories were quite unable to provide an intelligible scheme to account for the observed facts covered by Bohr's frequency condition and Ritz's combination law of spectroscopy. When faced with the task of accounting for such observational generalizations, classical theories degenerated into a mass of *ad hoc* postulates, assumptions, and rules of thumb which enjoyed neither intrinsic plausibility nor generality of application. It has been the peculiar merit of the newer forms of quantum theory that they have succeeded, where classical theory failed, in establishing a set of principles susceptible of fruitful mathematical development, enabling the experimental facts to be systematized in an elegant, consistent and intelligible account or picture.

But although there is general agreement regarding the actual *formal expression* of the fundamental laws of the new quantum theories, there are, as everybody knows, several different ways of approaching and interpreting these formal expressions, each leading to what may loosely be called a "picture" of a rather different kind.

Dirac, in 'Quantum Mechanics,' sought to unify these diversities of approach and to co-ordinate the different quantum "pictures" in a

* Communicated by Professor G. H. Hardy, F.R.S.

single comprehensive formal structure rigorously developed from a certain set of initial assumptions; and many might be tempted to argue that his apparent success in this task was one of his principal achievements. To say this, however, is in no way to belittle the very important specific contributions which he has made from time to time to the general development of quantum theory. Among these I need only mention his formulation of a relativistic wave equation for the electron, which, despite its apparent lack of symmetry, can be shown to be invariant under a Lorentz transformation. Also, his highly original and suggestive theory of the positron and his use of sets of anticommuting Hermitian operators to represent the spin variables associated with the electron. But to many, I think, the most pleasing feature of his work in 'Quantum Mechanics' is his apparent success in unifying the discrete and continuous by means of an elegant and comprehensive general representation theory, in which a close formal parallelism is maintained between the two. A good example of this can be seen in Dirac's formulation of the principle of second quantization which (as I shall later show) depends on his general representation theory. This representation theory rests ultimately upon a peculiar and characteristic view of the relationship between discrete and continuous point sets, which I shall refer to as Dirac's assimilation postulate. This principle colours his whole approach to the problems of quantum theory. Now, on the strength of his assimilation postulate, Dirac develops an elegant and highly unified picture of quantum phenomena, in which we can see the exact relationship of the older classical theories of mechanics and the new quantum views clearly portrayed. But it is clearly impossible within the compass of a note of this kind to present Dirac's theory in its full deductive rigour. In what follows, therefore, I shall have to confine my attention to those parts of the general theory to which I shall specially wish to refer later on when discussing the paradox and its origins. With this object in view I shall pick out a certain fundamental set of assumptions, theorems and propositions occurring in Dirac's representation theory, and I shall number these consecutively to facilitate reference to them later on in the note. But, in order to make reference back to Dirac's original work as easy as possible for the reader who is sufficiently interested, I propose in general to adopt Dirac's notation; and I shall quote the corresponding chapter and equation reference number in 'Quantum Mechanics' (second edition) wherever possible, putting this reference within a square bracket to distinguish it from *my* equation numbering, which will be found in a round bracket immediately to the left of it. I hope this procedure will be found satisfactory; it appears to me to be the only practicable way of bringing out the point I have in mind within the limits of a short article. Anyone who is concerned to carry the work a stage further will of course have to check my conclusions by careful reference to Dirac's own work.

I shall begin by explaining one important sense in which Dirac uses the word "state" as applied to a dynamical system in quantum theory.

Consider a general dynamical system composed of particles moving and interacting according to certain specified laws. These laws will permit various possible motions of the particles, and each such motion is called a "state" of the system. According to classical ideas, it would theoretically be possible to specify a "state" with complete determinateness by assigning numerical values simultaneously to all the relevant variables. For this picture quantum theory substitutes the notion of a state which is describable in terms of motions restricted by as many conditions as are theoretically possible. The general quantum principle of superposition then applies to the states of this kind belonging to any one dynamical system. The essence of the principle is the assumption that a particular sort of relation holds between states, in virtue of which it is always possible to regard a system which is in one state as the result of superposing two or more other states, so that in a special sense this state may be said to be partly in each of the superposed states. Not only may any state, according to quantum theory, be regarded as the result of the superposition of other states, but also the converse is true, and any two (or more) states *belonging to the same dynamical system* may be superposed to give a fresh state. The superposition process is essentially a mathematical one, which can be effected (like the resolution of a wave into its Fourier components) irrespective of the particular *physical* conditions of the problem. A state which is formed from the superposition of two other states will have certain deducible relations to the superposed states, which will depend in part upon the relative "weights" of the two superposed states in the process of superposition. In fact, the resultant state will be specified as completely as is theoretically possible when the relative weights and phase difference of the states taking part in the superposing process are given. (The occurrence of this "phase" factor in the superposition is due to the fact that the vectors which represent the states being superposed can be multiplied by any complex number of modulus unity without altering either their length or direction.)

Since the superposition process is to be an additive one, holding between directed quantities, we may express these states of a dynamical system by means of vectors, and we can reduce the properties of the superposition operation to the following propositions:—

If the states A, B and C are represented by ψ_a, ψ_b, ψ_c , then if the state A can be formed by superposition from B and C we assume that

$$\psi_a = x_b \psi_b + x_c \psi_c, \quad \dots \quad (1) \quad [\text{D. Ch. I. (1)}]$$

where x_b and x_c are numerical coefficients. Here the relation of superposition is to be symmetrical between the terms. More generally, where we have a symmetrical relation of superposition holding between a number of states A B C . . . N, we assume that there is a vectorial relation of the form

$$x_a \psi_a + x_b \psi_b + \dots + x_n \psi_n = 0, \quad \dots \quad (2) \quad [\text{D. Ch. I. (2)}]$$

where not all the coefficients $x, x_b \dots x_n$ are zero.

Quantum theory also assumes that a state A may be superposed upon itself to yield precisely the same state A, *i. e.*, that

$$x_{a_1}\psi_a + x_{a_2}\psi_a = (x_{a_1} + x_{a_2})\psi_a, \quad . \quad . \quad . \quad (3) \quad [\text{Ch. I.}]$$

where x_{a_1} and x_{a_2} are numbers. That is to say, in the language of the theory of groups, the operation of superposition is to be idempotent. This idempotency is a most important property which differentiates the quantum from the classical idea of superposition. In a classical system superposition is not idempotent, and the result of superposing a state upon itself will be (*e. g.* in the case of a vibrating membrane) a state with oscillations of a different amplitude. Further, we can deduce from (3) that it is possible to multiply any vector ψ representing a non-classical state by any number other than zero, and the resulting vector will represent the same state.

Dirac's theory also assumes that complex numbers can occur as coefficients in superposition equations such as (1), (2) and (3) above. This assumption is made plausible by general physical considerations. For example, the states of elliptic polarization of a photon require in general two parameters for their specification, and consequently the totality of such states will be a two-fold infinity. Now we want in quantum theory to be able to regard all such states of elliptic polarization as derivable from the applications of a superposition process to two independent states of linear polarization (*viz.* parallel and perpendicular to the optical axis) with varying weights. But such superpositions would (if restricted to real coefficients) yield only a simple infinity of states, and the requisite two-fold infinity will only be forthcoming upon the assumption that the coefficients in the superposition equations are in general complex numbers.

The assumption of the possibility of complex coefficients has two important consequences :

- (1) In *every* case of superposition of two different states a two-fold and not merely a simple infinity of states is obtainable.
- (2) The vectors representing the states are in general complex vectors, and cannot be resolved into real and pure imaginary parts ; and it is clear that it will not be possible to give any invariant meaning to the notion of a real vector representing a dynamical state in Dirac's theory.

This conclusion has important consequences. For, consider a vector ψ_a whose co-ordinates in some simple system are a set of numbers a_s . These co-ordinates will generally be complex, since it follows from the assumptions just made that the vector ψ_a may be multiplied by a complex coefficient and will still represent the same state. Now consider a co-ordinate transformation to another simple system in which the vector ψ_a is represented by another set of numbers a_r . It is clear from our

assumptions that such a transformation can be represented by any linear expression of the form

$$a_r' = \sum_s \gamma_{rs} a_s, \quad \dots \quad (4) \quad [\text{Ch. II. (1)}]$$

where the γ_{rs} are numbers which may be complex.

Since in general any vector representing a dynamical state will be complex and have complex co-ordinates, a_s say, we may consider the conjugates of these a_s . Now Dirac assumes that these will be the co-ordinates of another vector situated in a different vector space. This he denotes by the symbol ϕ_a , so that the co-ordinates of ϕ_a are conjugate complex to those of ψ_a . They will be transformable according to linear laws of the form

$$a_r' = \sum_s \overline{\gamma_{rs}} a_s, \quad \dots \quad (5) \quad [\text{Ch. II. (2)}]$$

where the bars denote the operation of taking the conjugate complex of the number beneath.

These two kinds of vector are not in the same space, so that a formula such as $\phi_a + \psi_a$ has no meaning. But the two different vector spaces they inhabit have definable relations so that, for one thing, the product of a ϕ - and a ψ -vector is invariant under a linear co-ordinate transformation. This suggests an analogy with the covariant and contravariant vectors familiar in the differential geometry of Levi-Civita. Then the product $\phi_a \psi_b$ in Dirac's theory corresponds to the inner or contracted product of two vectors A_μ and B^ν , and multiplication of any vector ϕ_a by its conjugated imaginary ψ_a is analogous to the process of deriving an invariant by the operation of contraction in tensor analysis.

The next step in Dirac's vector theory is the idea of representing observables (the dynamical variables which are directly measurable) by real linear operators, *i. e.* those operators often called Hermitian, whose elements are such that $\alpha_{rs} = \overline{\alpha_{sr}}$. He points out that this assumption is equivalent to making the representative of an observable analogous to a real function of the dynamical variables, whereas representation by a general linear operator would correspond rather to a complex function. The justification pleaded by Dirac for imposing this restriction cannot be said to be very plausibly stated by him. It seems to come to saying no more than that an observable must itself be real in order to be observed at all, and therefore, if it is to be a linear operator, it must be a real (*i. e.* a Hermitian) operator. I cannot see that there is anything in this. For a similar argument would surely apply to the representatives of the dynamical states. For it might be argued, with at any rate equal plausibility, that these too must be real to exist; and must presumably be every bit as real as the observables with which they are associated, if the system is to be of the kind that could actually exist. Thus, from an ontological point of view, it might well be argued that both states and observables are on a precisely similar footing, so that it would be an entirely false and artificial distinction to infer from purely ontological considerations that observables *must* be represented by real numbers

whereas the states associated with them might be represented by complex co-ordinates.

In fact, I believe that the real reason which has swayed Dirac's decision in this matter is quite different from the one he has actually given and is a perfectly genuine one needing no support on flimsy ontological grounds. This will, I think, become more obvious if we consider what I shall call his fundamental eigen theorem, which he introduces with these words on page 30 —

"In the special case when the result of a particular observation made on the system in a particular state is with certainty one particular number, 'a' say (instead of being one of two or more numbers according to a probability law), then the Hermitian operator representing the observable that is measured and the ψ -vector, ψ_a say, representing the state are connected by the equation

$$\alpha\psi_a = a\psi_a." \quad . \quad . \quad (5) \quad (\text{D. Ch. II. (12)})$$

He also says a little further on: "There are some other matters we must look into before we can be sure our assumption (12) is reasonable. One of these concerns the reality of the number 'a.' *Any result of a measurement is necessarily a real number.* Is any number 'a' satisfying an equation of the type above also necessarily real? We can easily see that it is so when we make use of the Hermitian property of α ."

The words I have italicised in this quotation give the real reason for Dirac's restriction of the representatives of observables to Hermitian operators. But the assumption "Any result of a measurement is necessarily a real number" needs justification, and a completely adequate one could only be given by a thorough-going examination of the theory of measurement. This I cannot attempt here for obvious reasons of space. But since this restriction imposed by Dirac on the representatives of observables affects the whole subsequent course of his theory, it is of peculiar importance; and it would be decidedly unfortunate if it turned out to be founded on nothing more solid than a dubious philosophical prejudice. Now I do not for a moment believe this to be the case, but since I shall be concerned later to cast doubt upon certain other fundamental assumptions made by Dirac, I think it behoves me to indicate very briefly the sort of reasons which do, I think, provide some justification for the assumptions on which the validity of the fundamental eigen-value theorem turns, even at the risk of some digression from my main argument.

The true justification of this assumption may, I think, be found in the conventional features which permeate all systems of measurement placed on a properly relativistic basis. For, according to general relativistic requirements, in the actual world only those events which are separated from observation events by intervals not intersecting the light cone can be observed. In other words, anything which is an observable must be associated with a world line which does not at any point pass within the null cone, so that the region of space-time which is available for the loca-

tion of observables is the part lying outside the null cone. Now it can easily be shown that the invariant quadric $(G_{\mu\nu} - \frac{1}{2}g_{\mu\nu}G)dx_\mu dx_\nu = 3$ must be real for intervals between events lying in this region. This quadric is a measure of the curvature of the continuum, for its radius is equal to the radius of spherical curvature of the corresponding 3-dimensional section of space-time. So we have the result that for the region of space-time in which observables can occur the directed radius of spherical curvature of a 3-dimensional section of the world is given equal to the radius of a real quadric. Now a most fruitful and illuminating way of looking at Einstein's law of gravitation, $G_{\mu\nu} = \lambda g_{\mu\nu}$, is to regard it as equivalent to the assertion that the length of a standard material structure bears a constant ratio to the radius of spherical curvature at the place and in the direction in which it lies. But all actual measurements made upon observables imply in the last analysis a reference to lengths measured by just such an ultimate standard. Thus we have the necessary result that all such measurements must be represented by a set of coefficients which bear a constant ratio to the radius of a real invariant quadric, and that is to say they must be represented by *real* numbers. Moreover, this result, since it is expressed in tensor form, must be quite independent of the particular co-ordinate systems which may be used for the specification of the physical continuum. For, no matter what particular co-ordinates are used, the quadric of curvature for the regions open to observation must remain invariant and real; and this is just the result needed to justify Dirac's assumption in the formulation of his fundamental eigen-theorem to which I have already referred. Further, although the rationale I have suggested for this assumption only applies, in the way I have expressed it, to a continuum obeying the Riemannian geometry presupposed in Einstein's exposition of the general theory of relativity, yet it can be generalized so as to cover all cases of measurements, made by both optical and material methods, for manifolds with non-Riemannian geometries which permit of unlimited measure gauge transformations. In the general case of manifolds having such non-Riemannian geometries, it is only necessary to develop more general tensors (which remain invariant under unlimited gauge transformations) to take the place of Einstein's expressions (since though these are invariant for *co-ordinate* transformations they are not so in general for the unlimited gauge transformations allowed by the generalized geometries). It seems reasonable to assume that the natural gauge metric system of the world is one which makes it a self-gauging unit, since any measuring appliances are parts of the world and thus we must suppose that the natural gauge-system is intrinsic to its structure. This self-gauging character of the world suggests that to formulate or express our natural gauge-system we should look for suitable fundamental gauge invariant tensors, simply and directly derivable from the basic expressions used to define the (as yet) non-metrical geometrical structure of the continuum. Eddington has shown that a gauge invariant tensor of 2nd rank $*G_{\mu\nu}$ can be found having the properties required in a continuum

with a perfectly general relation structure, limited only by what he calls the general parallelogram law, namely, that

$$\Gamma_{\nu\alpha}^{\mu} = \Gamma_{\alpha\nu}^{\mu},$$

where the $\Gamma_{\nu\alpha}^{\mu}$ represent a 3-fold array of 64 terms involved in the expression for a generalized differential parallelism. This tensor $*G_{\mu\nu}$ is derived by contraction from the more fundamental gauge invariant tensor of the 4th rank $*B_{\mu\nu\sigma}^{\epsilon}$, which appears as a factor in the integrand of the expression for the change produced in a vector A^{μ} carried round an infinitesimal circuit by a parallel displacement obeying the parallelogram law. Eddington then suggests the formula

$$l^2 = *G_{\mu\nu} A^{\mu} A^{\nu} \dots \dots \dots (6)$$

as the expression for the natural gauge of the world. In this formula the antisymmetrical part of the tensor $*G_{\mu\nu}$ drops out, leaving a symmetrical part which is the generalized analogue of Einstein's $G_{\mu\nu}$, and considerations similar to those advanced above in connexion with Einstein's formula lead to the same conclusion in the generalized case, namely that actual measurements must be expressible in terms of *real* numbers.

We may conclude, therefore, that the assumption contained in Dirac's eigen-value theorem is justifiable; and it was, I think, necessary to digress to make ourselves sure of this because of the fundamental importance of this proposition to the whole of his theory. In this connexion we may quote his own words (page 32 of Chapter II.): "Equation (12) of Ch. II. is of such fundamental importance in the theory that it is desirable to introduce some special words to describe the relationship between the quantities involved. We shall call a an 'eigen value' of the operator α or of the observable that α represents and ψ_a an 'eigen- ψ ' of this operator or observable, and we shall say that the eigen- ψ ψ_a belongs to the eigen value a . Likewise ϕ_a satisfying $\phi_a \alpha = a \phi_a$ is an eigen- ϕ belonging to the eigen value a , and the state represented by either ψ_a or ϕ_a is an *eigen state* belonging to the eigen value. This terminology may also be used when the linear operator α is not Hermitian and does not represent an observable.

"Our assumption now enables us to infer that every eigen value of an observable is a possible result of the measurement of that observable. It is certainly the result when the system is in an eigen state belonging to the eigen value."

After introducing this special terminology Dirac proceeds to develop certain useful basic properties of observables and their eigen values and eigen- ψ 's. Amongst other things he shows that the algebra of observables and Hermitian operators differs only from ordinary algebra in that the commutative axiom of multiplication does not hold in general. This, and other important properties belonging to observables and the operators which represent them, are of course the direct consequence of restricting these representatives to real linear operators so that they obey the normal algebra applicable to all Hermitian matrices. We must remember that the elements of the matrix correspond to the co-ordinates of the

operator, and that the condition for any matrix to be Hermitian is that the diagonal elements must be real and the corresponding elements on each side conjugate complex, *e. g.*

$$\begin{pmatrix} a_{11} & a_{12}+ib_{12} \\ a_{21}-ib_{21} & a_{22} \end{pmatrix} \quad (a_{12}=a_{21}, \quad b_{12}=b_{21}).$$

However, in the particular case of a real number, which may be regarded as a special case of a Hermitian matrix with just one eigen value, namely the number itself, the commutative law does of course hold in that a real number commutes with every linear operator. He also proves certain very important theorems connected with the property of orthogonality. These are :—

Two eigen states belonging to two different eigen values of an observable are orthogonal, *i. e.*

$$\phi_s \psi_s = 0. \quad (7)$$

If ψ_1 and ψ_2 are two eigen- ψ 's belonging to the same eigen value, then any linear combination of them of the form $c_1\psi_1+c_2\psi_2$ is also an eigen- ψ of this eigen value. But no linear combination of eigen- ψ 's belonging to different eigen values can be an eigen- ψ , so that any two eigen- ψ 's belonging to different eigen values are necessarily *independent*.

Up to this point Dirac has not discussed the distinction between the case of an observable whose eigen values consist of all numbers within a certain continuous range or interval, and that of an observable represented by an operator with a discrete enumerable number of eigen values.

In fact, this distinction has been unnecessary at this stage, as the discussion has been purely general and applies equally to both cases. But when he comes to develop the notion of the expansion theorem (which is fundamental to his general theory of representation) this distinction is forced upon his attention by certain difficulties which arise in the case of observables with a continuous range of eigen values.

This idea of an expansion theorem for discrete observables may be developed in the following way. Consider a Hermitian operator with co-ordinates given in some simple co-ordinate system in a vector space of a finite number of dimensions ; then the condition for the existence of an eigen value of a Hermitian operator α is

$$\sum_s \alpha_{rs} x_s = \alpha x_r, \quad (8)$$

where α is a real number and an eigen value of α . Then, if n is the number of dimensions of the vector space, we may eliminate the n variables x_r (which occur linearly and homogeneously), and can at once write down the condition that α is an eigen value of α by the vanishing of the determinant

$$\begin{vmatrix} \alpha_{11}-\alpha & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22}-\alpha & \dots & \alpha_{2n} \\ \alpha_{31} & \alpha_{32} & \dots & \alpha_{3n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn}-\alpha \end{vmatrix} = 0.$$

The determinantal equation for a is one of the n -th degree, having therefore n roots, some of which may coincide. Then the orthogonality theorem can be used in conjunction with the Hermitian character of α to show that, when the number of dimensions of the ψ space is finite and equal to n , the number of independent eigen- ψ 's of any Hermitian operator is also n , so that when two or more roots of the determinantal equation for a coincide at some particular value a_1 (say), then there must be as many independent eigen- ψ 's belonging to this value as there are roots coincident at it.

This result permits us to express any arbitrary ψ linearly, in terms of these independent eigen- ψ 's of a Hermitian operator representing an observable with a finite number of eigen values, by the equation

$$\psi = \sum \psi_a,$$

where each ψ_a is an independent eigen- ψ of α . This is the expansion theorem for a vector space with an enumerable number of dimensions. The case of an observable with a continuous range of eigen values is more difficult. Here we obviously cannot express the connexion between eigen- ψ 's and eigen values of a Hermitian operator, and the observable it represents, in the simple form of a sum such as $\sum_s \alpha_s x_s = \alpha x$. For the number of elements in each row of the (Hermitian) matrix will belong to a non-enumerable order of infinity.

Nor can the expansion theorem for the case of an observable with a continuous range of eigen values be expressed in general by an equation of the form $\psi = \sum \psi_a$, since the number of independent ψ 's will also be a non-enumerable infinity.

Dirac rides cheerfully over this difficulty by assuming that in such cases an expansion for any arbitrary ψ in terms of the eigen- ψ 's of a continuous observable can always be given the form $\psi = \int \psi_a da$, where ψ_a is an eigen- ψ belonging to the eigen value a and the schedule of eigen- ψ 's is so related to that of the eigen values that ψ_a may be regarded (at any rate from an extensional point of view) as a function of the variable " a ," which varies sufficiently continuously to permit of the existence of the integral $\int \psi_a da$ within the range considered. Dirac admits that no rigorous proof is forthcoming in support of the expansion theorem for the case of Hermitian operators with a continuous range of eigen values, but instead of feeling doubtful about the applicability of the expansion theorem to the notion of a continuous observable, he proceeds calmly to cut the Gordian knot by *postulating* that only those operators which do *in fact* obey the expansion theorem can represent observables. He tries to make this rather facile procedure plausible by making the (apparently) innocuous assumption that it will always be possible to expand an arbitrary vector ψ (which is a function of a parameter t) so that it can be expressed by Fourier analysis in terms of a Fourier series or integral. On this assumption it is easy to show that the arbitrary vector ψ will always be expressible in the form required

by the expansion theorem, namely either as a sum or as an integral of the form

$$\psi = \int \psi_a da.$$

But this procedure of Dirac's leaves much to be desired ; for we cannot be sure that it will always be the case that any arbitrary ψ representing a dynamical state is expressible as a Fourier series or integral. However, Dirac's representation theory requires an expansion theorem both for continuous eigen values and for discrete eigen values. This will become apparent when I come to discuss the relations between the representation theory for the two cases, and his attempt to assimilate the formal treatment of the two by means of certain rules of substitution based on the notion of an improper δ -function. But before I come to this stage I must first describe his representation theory for the two cases of discrete and continuous eigen values. I shall begin by confining my attention to the discrete case since, once the more important principles applying to this have been stated, the transition to Dirac's theory for the continuous case can be made quite simply on the basis of the principle which he puts forward on page 71 in these words :—"From general physical grounds, and from the possibility of regarding a continuous range of numbers as a limiting form of a discrete set whose density is increased indefinitely, one would expect the theory (*i. e.* the theory of representation for discrete and continuous eigen values) to run on somewhat parallel lines in the two cases."

Dirac initiates his account (which I shall have to condense as much as possible) of the representation theory for observables with a discrete range of eigen values by introducing the notion of a set of ψ 's (which he calls "basic ψ 's") chosen so as to have the following properties :—

- (1) They are all orthogonal to each other.
- (2) Each of them is normalized, *i. e.* has a length equal to unity according to a "natural" standard gauge-system (the precise significance of which will be made apparent in Part III. of this note).
- (3) There are enough of them to make any arbitrary ψ dependent on them, so that the number of basic ψ 's is equal to the number of dimensions of the vector space (which is of course enumerable in the case of representations for discrete eigen values).

Then he says : "Such a set of ψ 's will be called a set of basic ψ 's for a co-ordinate system. The co-ordinates of any ψ will then be its coefficients when expanded in terms of the basic ψ 's. We shall denote a co-ordinate thus associated with a basic ψ , ψ_r , by the bracket expression r)." ."

Similarly, the set of ϕ 's which are the conjugate imaginaries of the basic ψ 's of a representation will be the basic ϕ 's of a co-ordinate system in the ϕ -space, and Dirac writes the co-ordinate in the expansion formula

for an arbitrary ϕ associated with some basic $\phi_1\phi_r$ by the corresponding conjugate bracket expression ($|r\rangle$).

Using this notation the conditions for orthogonality and normalization for a set of basic ψ 's in a representation for an observable having a discrete range of eigen values can be written

$$\phi_r\psi_s = \delta_{rs},$$

δ_{rs} being Dirac's δ -symbol having the meaning

$$\delta_{rs} = 0 (r \neq s), \quad \delta_{rs} = 1 (r = s);$$

also an arbitrary ψ , ψ_x , may be expanded in terms of basic ψ 's as :

$$\psi_x = \sum_r \psi_r(r|).$$

The next step in the development of the discrete representation theory is to take the eigen- ψ of some suitable observable in the system as the basic ψ of the representation. If there are in fact two or more eigen- ψ 's belonging to the same eigen value of the observable (which will be represented by the Hermitian operator α say), then it is necessary to select as basic ψ 's a subset from the set of coincident eigen- ψ 's each member of which is normalized and orthogonal to every other basic eigen- ψ chosen to be a basic ψ . It is then easy to show that the co-ordinates of the observable represented by the operator α itself are just the diagonal elements of the Hermitian matrix α , i. e. those elements of the form α_{11} , α_{22} , α_{33} . . . , α_{nn} which are the eigen values of α . A representation of this kind Dirac calls a diagonal representation. Its arbitrariness (arising from the possibility of a multiplicity of coincident eigen- ψ 's) as a frame of reference can be eliminated by introducing a set of additional observables each of which commutes with α and every other member of the set, and is sufficiently numerous to ensure that there is never more than one eigen- ψ which belongs simultaneously to eigen values of each member of the set of commuting observables. Then a representation can be set up in which each member of the set of commuting observables is represented by the diagonal elements of the corresponding Hermitian matrix, and such a set is called by Dirac "a complete diagonal representation." *À propos* of this he says, on page 58, "This kind of representation is the most useful one in practice. In it each of the complete sets of commuting observables will be diagonal. Further the representation will be completely determined by the complete set of commuting observables except for arbitrary phase factors arising from the fact that the basic ψ 's may be multiplied by arbitrary numbers of modulus unity (i. e. numbers of the form e^{ic} , where c is a real number) without any of the conditions defining them being invalidated." Let us now represent the first complete set of commuting observables, $\xi_1, \xi_2, \xi_3, \dots, \xi_m$ by Dirac's symbol ξ_m , and then consider the existence of another complete set of commuting observables $\eta_1, \eta_2, \eta_3, \dots, \eta_l$ (which we may write η_l for short in Dirac's notation). Clearly it will be possible to set up another diagonal representation for this second set.

This representation will differ from that in which the ξ_m are diagonal. But there will be definite transformation relations between the two, since this follows from the passage we have just quoted expressing the determinateness of such diagonal representations. Thus any arbitrary ψ will now have *two* representatives, one in the ξ_m representation and the other in the η representation. These will be in Dirac's notation

$$\psi = \sum_{\xi'} \psi(\xi') (\xi' |), \quad \dots \quad (9) \quad [\text{Ch. III. (31)}]$$

and

$$\psi = \sum_{\eta'} \psi(\eta') (\eta' |), \quad \dots \quad (10) \quad [\text{Ch. III. (32)}]$$

where $(\xi' |)$ and $\psi(\xi')$ are compendious symbols denoting co-ordinates and the basic ψ 's associated with them in the ξ_m representation, and similarly with $(\eta' |)$ and $\psi(\eta')$ in the η_m representation.

Now each basic ψ of the η -representation will itself have a representative in the ξ -representation. This ξ -representative of the basic ψ $\psi(\eta')$ Dirac writes $(\xi' | \eta')$. Similarly, each basic ψ $\psi(\xi')$ of the ξ -representation will have an η -representative which Dirac writes $(\eta' | \xi')$. Now we have

$$\psi(\eta') = \sum_{\xi'} \psi(\xi') (\xi' | \eta'). \quad \dots \quad (11) \quad [(33)]$$

Thus the expression for an arbitrary ψ can be written

$$\psi = \sum_{\xi'} \psi(\xi') (\xi' | \eta') (\eta' |), \quad \dots \quad (12) \quad [\text{Ch. III. (34)}]$$

which gives

$$(\xi' |) = \sum_{\eta'} (\xi' | \eta') (\eta' |). \quad \dots \quad (13) \quad [(35)]$$

Similarly, we may obtain

$$(\eta' |) = \sum_{\xi'} (\eta' | \xi') (\xi' |). \quad \dots \quad (14) \quad [(36)]$$

These two latter expressions are just the transformation equations which give the representatives of two diagonal representations for two distinct and complete sets of commuting observables in terms of each other for the case of observables with a discrete range of eigen values.

Further, it can easily be shown that consistency requires that

$$\sum_{\eta'} (\xi' | \eta') (\eta' | \xi'') = \delta_{\xi' \xi''}$$

and

$$\sum_{\xi'} (\eta' | \xi') (\xi' | \xi'') = \delta_{\eta' \eta''},$$

where the δ -symbol has the usual significance attached to it in Dirac's notation, namely

$$\delta_{\xi' \xi''} = 0 \quad (\xi' \neq \xi''), \quad \delta_{\xi' \xi''} = 1 \quad (\xi' = \xi'').$$

In such transformations there will be an indeterminacy of phase, since if the basic ψ 's $\psi(\xi')$ and $\psi(\eta')$ are multiplied by expressions of the form $e^{if(\xi')}$ and $e^{ig(\eta')}$, where " f " and " g " are real functions of the variables ξ' and η' , the transformation function will have a factor equal to $e^{-i[f(\xi') - g(\eta')]}$.

But since " f " and " g " are both real functions this expression must be of unit modulus, so that it affects only the phase of the transformation and not the value of the modulus, the latter being perfectly determinate. This completes all that I shall say by way of introduction to Dirac's

theory of representation for observables having a discrete range of eigen values, and I must now go on to consider treatment of continuous eigen values.

He says, lower down on page 72: "Let us take an observable ξ with a continuous range of eigen values and suppose for the present that it has only one independent eigen- ψ belonging to any eigen value. Then, ignoring for the present the question of normalization, we can take its eigen- ψ 's, $\psi(\xi')$, as the basic ψ 's of a representation. The number of these basic ψ 's, equal to the number of axes of our system of co-ordinates, is an infinity of a high order equal to the number of points on a line, but this is not in itself a source of difficulty. The fundamental equation defining the representative of a ψ must now read

$$\psi_x = \int \psi(\xi') d\xi'(\xi'|x), \quad . \quad . \quad . \quad [\text{Ch. IV. (1)}]$$

with an integral instead of a sum, the range of integration being understood to be the range of eigen values of ξ ."

Unfortunately for this attempt to develop the theory for continuous eigen values on lines analogous and parallel to the discrete theory, Dirac is faced with a serious difficulty at the start, namely the impossibility of expanding any basic ψ in the above form. If we compare the formula above with the corresponding one for the discrete case, namely $\psi_x = \sum_r \psi_r(r|x)$, the reason will become immediately obvious. For clearly this formula, when applied to the case of a basic ψ , $\psi(\xi'')$ say, reduces to

$$\psi(\xi'') = \sum_{\xi'} \psi(\xi') \delta_{\xi' \xi''}.$$

But we cannot produce anything analogous in the case of a continuous variable, *i. e.* where the observables have a continuous range of eigen values, because the symbol $\delta_{\xi' \xi''}$ has no meaning when it refers to a parameter with a continuous range of variation and is associated with an integration sign. To get round this initial, but quite fundamental, difficulty Dirac introduces his improper function $\delta(x)$, defined by

$$\left. \begin{aligned} \int_{-\infty}^{\infty} \delta(x) dx &= 1, \\ \delta(x) &= 0 \quad (x \neq 0). \end{aligned} \right\} . \quad . \quad . \quad (15) \quad [\text{Ch. IV. (2)}]$$

He says, on page 72: "Such an improper function may be pictured as the limit of a sequence of ordinary functions, corresponding to the fact that a ψ which cannot be expressed in the form (1) with $(\xi'|)$ an ordinary function of ξ' may be regarded as the limit of a sequence of ψ 's that can."

Of course, once having produced his improper function, Dirac does not bother any further over this fundamental difficulty of the continuous representation theory, but proceeds at once to deduce the required formula for a basic ψ in terms of the δ -function. This is

$$\psi(\xi'') = \int \psi(\xi') d\xi' \delta(\xi' - \xi''), \quad . \quad . \quad . \quad (16) \quad [(5)]$$

where $\psi(\xi')$ is a (continuous) function of ξ' .

After this initial but decisive step has been taken, the remaining properties required for a representation theory for observables having a continuous range of eigen values are easily worked out.

In particular, the conditions of orthogonality and normalization which the basic ψ 's must satisfy, in the case of continuous eigen values, may be expressed by

$$\phi(\xi')\psi(\xi'')=\delta(\xi'-\xi''). \quad (17) \quad [\text{Ch. IV. (21)}]$$

This formula corresponds to the condition

$$\phi(\xi')\psi(\xi'')=\delta_{\xi'\xi''}$$

for basic ψ 's in the case of discrete eigen values. But a basic ψ normalized according to the formula for the continuous case is not of unit length but of an infinite length. This follows from the character of the δ -function, which may be conceived as a quantity which is of such transinitely great magnitude that its integral over an infinitesimal domain in the neighbourhood of a point at the origin is unity. This consideration turns out to be of some importance when we seek to interpret certain formal expressions in the theory in terms of probability amplitudes. (In the case of the discrete theory the moduli of the co-ordinates of an arbitrary ψ are directly equal to probability amplitudes.) But otherwise it has no particular significance for the purely formal development of the parallelism between the representation theory for discrete and continuous eigen values. On the basis of these assumptions, and using a generalized conception of the notion of a "matrix" (making it appropriate to a two-dimensional array of a non-enumerable infinity of elements), Dirac concludes that "We can now see how the whole representation theory of the preceding chapter (dealing with representations for discrete eigen values) may be taken over to the case of a continuous range of basic (eigen) states. We simply have to replace sums by integrals and the two-suffix δ -symbol $\delta_{\xi'\xi''}$ by the δ -function $\delta(\xi'-\xi'')$ all the way through."

This happy conclusion greatly shortens and simplifies Dirac's task in providing a representation theory for the case of continuous eigen values. For on this basis, as he says (p. 81), "The transformation theory of § 17 (*i. e.* of the discrete eigen value case) may be taken over in the same way" There will then be transformation functions $(\xi'|\eta')$ and its conjugate complex $(\eta'|\xi')$, which enable one to pass from a ξ -representative to an η -representative by formulas analogous to those given above with sums replaced by integrals. The conditions which the transformation functions have to satisfy will be

$$\left. \begin{aligned} \int (\xi'|\eta')d\eta'(\eta'|\xi'') &= \delta(\xi'-\xi''), & (18) \\ \int (\eta'|\xi')d\xi'(\xi'|\eta'') &= \delta(\eta'-\eta''). & (19) \end{aligned} \right\} \quad [\text{Ch. IV. (30)}]$$

It would be quite possible for one representation to have continuous eigen values and the other discrete. (Dirac adds as a footnote to this

last statement : " If the number of basic ψ 's is finite in one representation it must of course be finite and equal in any other representation, but it may be infinite enumerable in one representation, and infinite equal to the number of points on a line in another.") We should then have similar transformation equations, with sums and integrals each occurring in the appropriate places. Instead of (30) we should have, if ξ' were continuous and η' discrete,

$$\left. \begin{aligned} \sum_{\eta'} (\xi' | \eta') (\eta' | \xi'') &= \delta(\xi' - \xi''), \quad (20) \\ \int (\eta' | \xi') d\xi' (\xi' | \eta'') &= \delta_{\eta' \eta''}. \quad (21) \end{aligned} \right\} \dots \dots \dots [(31)]$$

After this step two more steps are necessary on Dirac's theory to provide an entirely general representation theory. These are :

- (1) To allow some of the ξ 's in the complete commuting set to have discrete eigen values and others continuous eigen values, he says : " For each variable independently we must use either a sum or an integral, and either the two suffix δ -symbol or the δ -function according to whether it is discrete or continuous."
- (2) To allow any ξ to have as eigen values a discrete set of numbers together with a continuous range. This would give a representation theory in which sums and integrals occur added together in the formulas.

Such perfectly general representation can have transformation functions which relate them to other diagonal representations. These functions will be the corresponding generalized versions of the ones already given, but their detailed development and formulation need not detain us here.

It should now, I think, be clear that Dirac's representation theory essentially consists in selecting a set of eigen- ψ 's having certain definable properties in order to use them as the co-ordinate axes of a reference system. The object of this procedure is to obtain expressions for the fundamental quantum conditions which will be particularly appropriate to the requirements of theory on the one hand, and on the other hand which will permit of a simple and direct application to physical problems, by providing a straightforward interpretation of the co-ordinates of vectors representing dynamical states in terms of probability amplitudes. Thus, Dirac is in effect putting forward his complete diagonal representation as the " natural " co-ordinates for the expression of the fundamental quantum conditions. That this is his object is, I think, made abundantly clear by the following quotation with which he first introduces his procedure for taking eigen- ψ 's as the basic ψ 's of a representation (page 55) : " The representations we have used up to the present have all been quite general. We must now consider the question of how to introduce a particular representation which shall be advantageous for some special problem. . . , "

Moreover, *à propos* of the notion of a set of basic ψ 's (in the representation theory for discrete eigen values), he says (page 49): "... There are so many of them (*i. e.* basic ψ 's) that an arbitrary ψ is dependent on them, so that if the space has a finite number of dimensions there must be the same number of basic ψ 's." In connexion with the introduction of a representation theory for continuous eigen values he says (page 71): "The number of these basic ψ 's (afterwards identified with a set of eigen- ψ 's of an observable with a continuous range of eigen values), equal to the number of axes of our system of co-ordinates, is an infinity of a high order equal to the number of points on a line. . . ." Also, we must recall that Dirac says on page 15, "If we obtain the maximum number of independent states this will give us the number of dimensions of our vector space. In most practical examples this number is infinite. The vector picture is useful in spite of this, most of the reasoning that we use for it being equally applicable whether the number of dimensions is finite or infinite." Thus it is evident that both in the discrete and continuous cases the number of basic states of any diagonal representation must be equal to the number of dimensions of the vector space. This means that if we can set up two distinct representations, each with a different but complete set of commuting observables diagonal for the same system, not only must definite transformation relations exist connecting the two representations but, as in the case of more familiar co-ordinate systems, there must be an exact one-to-one correspondence between the co-ordinate axes of the alternative co-ordinate systems; and this, as we have just seen, entails in Dirac's representation theory that the sets of basic ψ 's of the two alternative complete diagonal representations must correspond one-to-one. Indeed, were it to be otherwise, the validity of the whole representation theory of Dirac would break down, since a lack of one-to-one correspondence between the two sets of basic ψ 's would involve a difference in the number of degrees of freedom in two equivalent co-ordinate systems referring to *one and the same* dynamical system. We can, I think, dismiss the possibility of a dimensional ambivalence of this kind as being entirely contradictory, and contrary to the whole conception of dimensionality, which is implicit in the analytic representation of manifolds. But then we must, as I have said, conclude that there is a one-to-one correspondence between the basic ψ 's of any two or more complete diagonal representations when transformation equations can be found which connect the representations in such a way that they can definitely be said to represent one and the same dynamical system. It is this conclusion which seems to me to lead inevitably to the paradox in Dirac's account of quantum theory, which I am now going to describe in Part II. of this note, since I have now finished all I need to say by way of introduction to Dirac's general representation theory and his main views on the relationship between representations for observables with discrete and continuous ranges of eigen values respectively.

PART II.

I think perhaps the best way of bringing out the paradox in Dirac's representation theory will be to start by considering a concrete example of it. Perhaps the best for my purpose is the case of the "harmonic oscillator," which Dirac discusses on page 133 of his 'Quantum Mechanics' (second edition). This is a particularly useful example, since its simplicity and freedom from irrelevant complexity of detail enable the underlying principles to be clearly grasped. In his discussion of the harmonic oscillator Dirac considers a simple system with only one pair of canonical co-ordinates as variables and, neglecting relativistic variation of mass with velocity, takes the Hamiltonian of the system to be

$$H = \frac{1}{2m}(p^2 + m^2 w^2 q^2), \quad \dots \quad (22) \quad [\text{VI. (47)}]$$

where m is the mass of the particle and $w/2\pi$ the frequency.

The wave equation for an oscillator with this Hamiltonian is

$$i\hbar \frac{d}{dt}(q'|) = \frac{1}{2m} \left\{ -\hbar^2 \frac{\partial^2}{\partial q'^2} + m^2 w^2 q'^2 \right\} (q'|), \quad \dots \quad (23)$$

where $(q'|)$ has its usual significance in Dirac's notation. Now those wave functions which represent the stationary states (*i. e.* states not varying with time and therefore represented by a ψ vector whose direction remains constant) are certain periodic solutions of the equation above.

In fact, they are those for which the operator $i\hbar \frac{d}{dt}$ becomes equivalent to the operation of multiplication by an energy eigen value H' . Consequently, such solutions must satisfy the equation

$$H'(q'|) = \frac{1}{2m} \left\{ -\hbar^2 \frac{\partial^2}{\partial q'^2} + m^2 w^2 q'^2 \right\} (q'|). \quad (24) \quad [\text{VI. (62)}]$$

Consequently, the general solution of this equation will furnish us with a relation between eigen values of the operator H' and those of q' , yielding a transformation function of the form $(q'|H')$, in Dirac's notation, connecting the q' and H' representations. But the q' -representation has a continuous range of basic states, since the observable q has a continuous range of eigen values. And the H' representation has only a discrete set of basic states belonging to the discrete though infinite set of eigen values of the variable H , which are $\frac{1}{2}\hbar w$, $\frac{3}{2}\hbar w$, $\frac{5}{2}\hbar w$, $\frac{7}{2}\hbar w \dots$ to infinity.

The paradox arises here from the necessary one-one correspondence, which we have seen must exist on Dirac's transformation theory, between the basic ψ 's of any two complete diagonal representations for a single dynamical system.

For in the course of the argument (dealing with the harmonic oscillator) Dirac has assumed that the Hamiltonian of the system represented by the operator H forms a *complete set of commuting observables*. Thus he

says, on page 134: "We now introduce a Heisenberg representation in which H is diagonal. We shall assume that H by itself forms a complete set of commuting observables, and its eigen values can therefore be used for labelling co-ordinates in the representation. The justification for this assumption is that it leads, as we shall see, without inconsistency, to definite representatives for q and p ." Thus, according to Dirac's Transformation Theory, the eigen functions $(q'|)$, corresponding to eigen values H' , which give a stationary state of H , represent transformation equations between a continuous and a discrete representation. And this, as we have seen at the end of Part I., entails that there is an exact one-one correspondence between the basic ψ 's of the two representations. But clearly the number of eigen values of any observables having a continuous range is non-enumerably infinite and is equal to the number of points on a line; consequently the number of basic ψ 's in such a representation will also be equal to a non-enumerable infinity on Dirac's assumptions. Thus our reasoning leads inevitably to the conclusion that the existence of such solutions of the wave equation for the harmonic oscillator will require, according to Dirac's representation theory, that a set of basic ψ 's, equal in number to the points on a line, can be put into one-one correspondence with an *enumerably* infinite set.

Now it is generally held that the number of points on a line, or of the terms in any *continuous* series, must be different from the number of terms in an *enumerably infinite* series. Indeed, Dirac himself tacitly admits this at the beginning of Chapter IV., where he says, on page 71: "the number of these basic ψ 's equal to the number of axes of our system of co-ordinates is an *infinity of a high order*, equal to the number of points on a line, but this is not in itself a source of difficulty." The parts I have italicised in this passage indicate clearly to my mind that Dirac, in his continuous Representation Theory, accepts at any rate that part of the theory of transfinite cardinals required by the theory of functions to give satisfactory differentia between strictly continuous and merely compact series. That is to say, he recognizes at least two of the different orders of cardinal infinity first distinguished in the theory of transfinite cardinals elaborated by Cantor. According to this theory of transfinite cardinals the cardinal number of terms in a progression and other similar enumerably infinite series is \aleph_0 (Aleph-zero), the least of all the transfinite cardinals. Whereas the cardinal number of terms in all series having the kind of continuity characteristic of the real number series is 2^{\aleph_0} (two to the power of Aleph-zero), and this number is necessarily greater than \aleph_0 itself. Now it is a fundamental postulate of the theory of functions that a set of real numbers, forming the domain of a real variable, can be put into one-one correspondence with the points on a line segment. Indeed, the domain of a single variable is frequently referred to as a "linear point set" or "linear aggregate." Thus Dirac's assertion that the number of basic ψ 's in a continuous representation "is equal to an infinity of a high order equal to the number of points on

a line " must be taken as equivalent to saying that in any such representation the number of basic ψ 's is equal to 2^{\aleph_0} , i. e. is a non-enumerable infinity.

But, according to Cantor's Theory of transfinite cardinals, the number 2^{\aleph_0} must necessarily be greater than \aleph_0 itself. At this point, however, some reader may feel tempted to observe that the greater part of Cantor's theory of transfinite numbers depends upon the validity of Zermelo's Axiom ; and that this latter axiom, which is equivalent to the assumption that the continuum can be well ordered, is very far from enjoying an unassailable security. There are, however, several things to be said on this score.

(1) While it may be true that, were Zermelo's axiom to be false, this would invalidate **many** of the distinctions which Cantor draws between different members of the infinite series of transfinite cardinals, each of which is greater than its predecessors, this would not invalidate the distinction in magnitude which Cantor draws between the fundamental infinite numbers, namely \aleph_0 and 2^{\aleph_0} .

(2) A considerable portion of existing mathematical procedure tacitly makes use of Zermelo's Axiom, though often it is not specifically invoked. Indeed, the theory of functions, and all those branches of mathematics which require a clear distinction to be drawn between enumerable and non-enumerable aggregates and ranges of values, make use of the distinction between 2^{\aleph_0} and \aleph_0 ; and Dirac's statement that the number of eigen values contained in a continuous range " is an infinity of a high order, equal to the number of points on a line," is merely a particular instance of this procedure. In any event it is clear enough from the words I have just quoted that, whether or not Dirac accepts Zermelo's principle, he certainly admits the possibility of assigning an order of magnitude to the infinity associated with a continuous range of values, which is to be different from that assignable to an enumerably infinite aggregate.

Moreover, Dirac's transformation theory requires, as we have seen, a correspondence between enumerably and non-enumerably infinite sets of basic ψ 's, and such a correspondence implies at least some form of comparability between two distinct orders of infinity, which would certainly be impossible on the assumption that Zermelo's Axiom were false. In any case it is certainly true that $2^{\aleph_0} > \aleph_0$, and consequently Dirac is faced with the paradox of asserting, in accordance with the requirements of his representation theory, that an aggregate having 2^{\aleph_0} terms can be put into one-one correspondence with one having \aleph_0 terms although $2^{\aleph_0} > \aleph_0$. And the property of reflexiveness (which according to Cantor's theory of transfinite numbers is common to all transfinite cardinals) does nothing to help avoid this paradox with which Dirac is faced. For while it is true that 2^{\aleph_0} and \aleph_0 can in a particular technical

sense both be said to be reflectable into a proper part of themselves, yet this does *not* entail that $2\aleph_0$ can be reflected into \aleph_0 . For to say that \aleph_0 can be put into one-one correspondence with a proper part of itself is no more than to give recognition to the truth of such propositions as

$$\aleph_0 + \aleph_0 = \aleph_0, \aleph_0 \cdot \aleph_0 = \aleph_0, \text{ etc.,}$$

which show that the ordinary rules of finite arithmetic break down in the case of transfinite numbers. The significance to be attached to the symbols of addition, multiplication, etc., and to the notion of part and whole, are radically different for infinite aggregates. The tendency to appeal to the property of reflexiveness as a solution of the dilemma thus rests upon a gross confusion, which is somewhat analogous to the confusion between the properties of "compactness" and strict "continuity" applied to series. For the series of points on a line cannot be put into one-one correspondence with any series having only an *enumerably infinite* number of terms, despite the fact that some discrete series (*e. g.* the series of rationals) can be re-arranged or re-ordered so as to be compact.

Now the harmonic oscillator is merely a particularly simple instance of a paradox which must occur in a whole range of cases in Dirac's theory. For the same situation will recur whenever we obtain an eigen function of an operator H , representing a discrete set of stationary states of a system, expressed in terms of variables with a continuous range. But equations of this type are very common in quantum-theory, and indeed the ability to develop such relationships has proved to be one of its chief merits and a fruitful means of accounting for such observational generalizations as Bohr's frequency condition. In fact, if we take any dynamical system for which a set of canonical co-ordinates q_r, p_s can be found, and in which the Hamiltonian of the system is expressible as a power series of the q 's and p 's, then it is easy to show that the equation for determining the eigen values of H becomes

$$H\left(q_r', -i\hbar \frac{\partial}{\partial q_s'}\right)(q'|) = H'.(q'|). \quad (25) \quad [\text{VI. (37)}]$$

Now this is a partial differential equation between the function $(q'|)$, whose domain is a continuous range, and an unknown eigen value of H which we may write H' . Now, if we are considering systems in which the stationary states of H are discrete, these eigen values will form a discrete (though infinite) range; and thus there will be an enumerable totality of particular solutions of the partial differential equation, each of which will furnish a particular eigen function $(q'|)$ corresponding to every distinct stationary eigen value of H . Thus the totality of all such solutions will, on Dirac's theory, furnish a general transformation function of the form $(q'|H')$ connecting a continuous q -diagonal representation with a discrete H -diagonal representation. Moreover, I think it is clear that this paradox is inherent in Dirac's theory of representations. For it is one of the great virtues of the new quantum theory that it has proved able to give a satisfactory account of the experimental fact that

the energy levels of the stationary states in an atomic system form a discrete series ; and to have shown just how the terms of this series are related to other quantities associated with the system—a connexion which was altogether inexplicable on previous theories of the atom. Now such connexions are expressible by certain particular solutions of the wave equation for the system, and they will take the general form of equations involving one-many relations between discrete and continuous ranges of values. For many of the characteristics of such dynamical systems will only be expressible by variables with a continuous range, so that in accounting for the observational facts, covered by empirical laws such as Bohr's frequency condition, by particular solutions of Schrodinger's equation, we must naturally expect to come across equations like (24) and (25) which we have just discussed above. In the ordinary way there is of course nothing in the least vicious or improper about such equations, and they are a commonplace of mathematics. It is, however, clear that Dirac could never hope to fit them into his representation theory unless he were prepared to regard them as transformation functions of some kind. Realizing this he appears to have set himself to work out suitable equations to cover this point when he is discussing the general principles of his transformation theory ; and I have already quoted an extract from this part of his theory, though I offered no comment upon it at that stage. Dirac does not, however, seem to have felt any qualms about the permissibility of equations of the form of my (20) and (21) which I have quoted above. On the contrary, he seems to have accepted, no doubt for the reason I have given, this development of his transformation theory without giving thought to the paradox contained within it. Thus he says in a footnote to page 82 : “ If the number of basic ψ 's is finite in one representation, it must, of course, be finite and equal in any other representation, *but it may be infinite enumerable in one representation and infinite equal to the number of points on a line in another.* ”

The parts of this quotation which I have italicised are, as we have seen, equivalent on Dirac's representation theory to saying that an infinite and continuous aggregate, with a non-enumerable number of terms, can be put into one-one correspondence with an aggregate having only an enumerable number of terms ; and this, as we have just shown, must be contradictory judged from Dirac's standpoint, since he himself admits the validity of a distinction between enumerable and non-enumerable orders of infinity.

We may then, I think, sum up this part of the discussion by saying that Dirac is faced with this dilemma :—

(1) Either his general representation theory is valid, in which case those particular solutions of the wave equation which represent the stationary states of the system must be invalid.

(2) Or the solutions of the wave equation, which determine the stationary states, are perfectly valid ; but they cannot be regarded as

transformation functions, as they must be on Dirac's representation theory, *i. e.* the representation theory breaks down at this point.

Now it is clear that we cannot afford to accept the first alternative of the dilemma, since to do so would be to forego some of the most striking and satisfactory achievements of the new quantum theory. On the other hand, if the second alternative is adopted, we must conclude that Dirac's theory as it stands contains contradictory elements. Before closing this note, therefore, I feel it is necessary to try and discover precisely at what points these contradictory elements find their way into Dirac's theory.

I believe that these defects are ultimately to be traced to Dirac's continuous representation theory and to certain general principles or assumptions underlying its formulation. These principles can be seen to be bound up with a more general and fundamental notion which I shall call the assimilation postulate, since it is under the guidance of this postulate that Dirac seeks to assimilate the formalism of his continuous representation theory to that of the discrete theory. Indeed, this fundamental assumption may be said to colour Dirac's attitude to the relation of discrete to continuous ranges in general, with far reaching results on the development of his theory as a whole. The assumption I am referring to as the assimilation postulate is the belief (which he expresses on page 71) in "... the possibility of regarding a continuous range of numbers as a limiting form of a discrete set whose density is increased indefinitely. ..."

It is not, I think, very difficult to see why Dirac adopts this postulate. For (A) it supplies him with the basis for developing a representation theory for discrete and continuous observables, with a parallelism between the two which is sufficiently exact and close to allow of his taking over the formalism of the one wholesale and applying it to the other (subject to certain simple rules of symbolic substitution). This result, besides being æsthetically pleasing, has the great practical merit of simplifying the labour involved in formulating and developing quantum theory. (B) An assimilation postulate of Dirac's kind would clearly be necessary if a weight function $s'd\xi'$ is to exist of the kind described by Dirac on pages 86 and 87 of his book. This weight function s' is to represent the density of the discrete points in a small but discrete interval from ξ' to $\xi' + d\xi'$ in the domain of a variable ξ (which may comprise one or more degrees of freedom). It is supposed to become infinite in the limit when the extent of the interval $d\xi'$ tends to zero. This weight function plays a very important part in Dirac's treatment of certain interaction problems by the procedure of second quantization, in which he finds it necessary to pass from representations involving discrete parameters to continuous ones, and *vice versa*. Indeed, as we shall see shortly, the assimilation postulate is an essential foundation for Dirac's formulation of the procedure of second quantization, which provides him with a means of reconciling the wave and particle pictures of phenomena according to quantum principles.

The adoption of the assimilation postulate may also be the explanation of Dirac's readiness to tolerate the paradox, discussed in Part II. of this note, which arises from his general transformation theory, since were the postulate to be true the number of points on a closed continuous interval would be an *enumerable* infinity.

[To be continued.]

LXXIX. *The Problem of the Keyboard Instrument.*—III.

By LL. S. LLOYD*.

[Received July 5, 1943.]

THE first of this series of three papers with the above title began with a cycle of fixed tones corresponding to vibrations with frequencies in the proportions :

$$1, \frac{9}{8}, \frac{5}{4}, \frac{4}{3}, \frac{3}{2}, \frac{5}{3}, \frac{15}{8}, 2;$$

and, from a brief review of some of the musical evidence, it was concluded that those physics text-books written towards the end of the nineteenth century which described this cycle, without qualification, as a musical scale were in error. For the musical scale is a flexible affair. So far as the physical tuning of *concorde*s is concerned the reason lies in the laws of arithmetic; 2, 3, and 5 being prime numbers. But not only have some notes of the scale to be *mutable* to produce true *concorde*s; others which form no part of the prevailing harmony are far from being fixed in intonation. Our cycle would produce perfectly tuned triads on the tonic (doh), the dominant (soh), and the subdominant (fah), and could thus be used for the two forms of Amen with which we are familiar in the Liturgy, and for some of their decorated equivalents. The musical use of the cycle ends there. It breaks down, as a musical scale, in the second concord of the Reading rota "Sumer is icumen in," to which the date c. 1240 is now attributed; for that concord demands a perfect fifth between the second and sixth notes of the cycle.

We saw that Robert Smith found in such musical facts the necessity for temperament in the tuning of any keyboard instrument; and though his example might well have been followed by subsequent writers of text-books on sound, it is important not to be too critical of them. For their mistake arose from musical misconceptions which had a vogue in the nineteenth century as musical theory.

The criticism of the cycle offered by Robert Smith was concerned with only one of its defects, that which disqualified it for the music of his

* Communicated by Sir Edward Appleton, F.R.S.

own period—the eighteenth century. (The second edition of his ‘Harmonics’ was published a year before Handel died.) The cycle would impose a further limitation on the musical scale, one that is less obvious but goes deeper. If we take the series of frequencies of the cycle we find that each gives with the first one the ratio of either a major or a perfect interval: in turn they give a major (greater) tone, a major third, a perfect fourth, a perfect fifth, a major (greater) sixth, a major seventh, and an octave. The cycle, indeed, embodies this relationship to a tonic or keynote, without which it becomes musically meaningless. We may think of this relationship as expressed by the notes of the *tonic solfa* system, *doh, ray, me, fah, soh, lah, te, doh.*

But a tonic is not a necessary element of the structure of a musical scale, as is evident from the earlier (modal) music of Western Europe*, as well as from some modern music. To understand the scale of European music for the last five hundred years, we must think of it, not as a system of fixed *notes* bearing a relationship to a tonic, but as a fluid and well ordered community of *intervals*, a community which acquires, and in turn loses, certain patterns as classical tonality enters or leaves it at the will of composers. It is true that, to our modern ears, “Sumer is icumen in” sounds as if it were written in the key of F, mutable notes and all. But we must not assume that it was so heard by thirteenth-century ears. The danger of making any such assumption may be illustrated from the earliest compositions of the sixteenth century, which opened a new chapter in English music. For example, Taverner’s “Western Wynde” Mass, is written in the Dorian mode; though, in the technical language of the musician, it is in the Dorian mode transposed, as is shown by the single flat in what we should call the key signature. To our modern ears, the opening strain of this Mass seems to alternate between the keys of G minor and F major without modulations that “establish” either key, and occasionally to touch the relative major and minor of these two keys, B♭ major and D minor. The “melody” (as we should call it) of this opening strain runs, in Hadow’s phrase, like a gold thread through the whole work, now in the treble, now in the bass, at other times in the tenor. And as many have read of this famous Mass but without seeing a note of it, here is the opening strain (fig. 1)*.

Observe that the chromatically altered notes in this quotation are B♭, E♭, and F♯. The Dorian mode itself required a C♯ for an authentic cadence on D (*i. e.* a full close or Amen, such as concludes our quotation

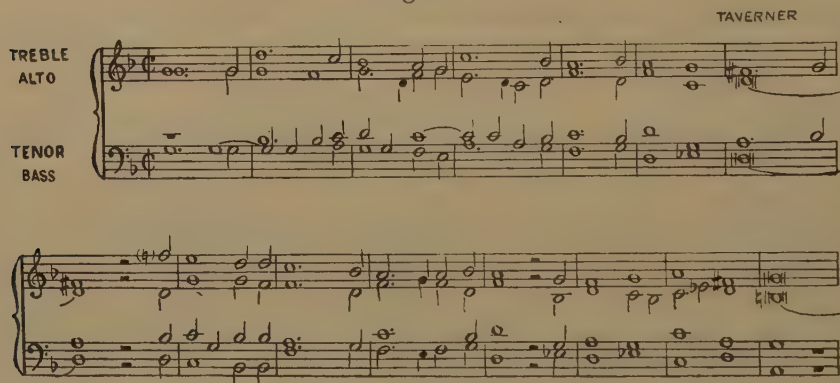
* To illustrate this Helmholtz quoted the opening phrase of Palestrina’s “Stabat Mater.”

* The writer is indebted to the Carnegie United Kingdom Trust for permission to transcribe this example from Vol. I. of ‘Tudor Church Music’ (he has inserted modern barring). Of Taverner, the editors of this monumental product of English musical scholarship write: “His mastery of a most intricate idiom and his amazing vitality and virility place him among the world’s great composers.”

on G), while an authentic cadence on A required a G#. With these additions this musical quotation will explain why, for a keyboard instrument, mean-tone temperament was the best available approximation to the flexible intonation of modal music. The noteworthy thing is that it was so acceptable for later music that its reign extended for some 250 years after its basis in modal music had disappeared.

As one accustoms oneself to modal compositions like this, one begins to recover, however imperfectly, the modal sense of music composed before tonality asserted the influence it exercised in the following two centuries or more. We must not think of a mode as something equivalent to a "key" as we know it to-day. Using many of the same notes and intervals, different modes gave them, as it were, a difference of atmosphere or general colour. But to those who are not acquainted with the music of the sixteenth century the difference of outlook may be more evident

Fig. 1.



from the diagram of the *gam*, or scale, with which Thomas Morley begins the instruction of his pupil in his "Plaine and Easie Introduction to Practicall Musicke" (1597). This immortal work has been made generally available in the reprint issued by the Shakespeare Association*, to whose kindness I am indebted for the two reproductions from its pages which follow. First Morley's diagram of the *gam* (fig. 2).

This diagram reveals clearly the hexachordal basis of the musical scale of the sixteenth century, using the six "notes" from *ut* to *la* to describe its community of intervals. The difference between this conception of the scale and that embodied in *tonic solfa* is very evident from an exercise in reading music which follows a few pages later in the 'Plaine and Easie Introduction' (fig. 3).

It is interesting to add that the musical nomenclature of the first example still held good in Robert Smith's day. For the frequency

* 'Shakespeare Association Facsimiles, No. 14' (Oxford University Press).

Fig. 2.

cc	la	la	1 note.
cd	la sol	sol la	2 notes.
cc	sol fa	fa sol	2 notes.
bb	fa mi	mi fa	2 notes, 2 cliffs.
aa	la mi re	re mi la	3 notes.
aa	sol re ut	ut re sol	3 notes.
e	fa ut	ut fa	2 notes.
d	la mi	mi la	2 notes.
c	la sol re	re sol la	3 notes.
b	sol fa ut	ut fa sol	3 notes.
a	fa mi	mi fa	2 notes, 2 cliffs.
G	la mi re	re mi la	3 notes.
F	sol re ut	ut re sol	3 notes.
E	fa ut	ut fa	2 notes.
D	la mi	mi la	2 notes.
C	sol re	re sol	2 notes.
B	fa ut	ut fa	2 notes.
A	re	re	1 note.
G	ut	ut	1 note.

Double or Treble keys. Same keys. G-clef or Bass keys.

Secunda ut prima. Sexta ut sexta. Quinta ut secunda. Quarta ut prima. Tercia deductio. Secunda deductio. Prima ex vocem deductio.

Fig. 3.

Phi.

sol sol la sol la fa sol fa mi sol la sol.

Ma. This is well song. Now here be diuerse other examples of plainfong, which you may sing by your selfe.

sol fa fa sol fa fa la sol sol fa la la sol fa mi la la sol.

fa sol la fa sol la sol fa fa sol la fa sol la fa sol sol fa.

fa sol sol la sol fa sol fa fa la fa sol fa fa mi fa.

sol la la sol fa la fa sol la sol sol fa sol la la sol.

sol fa mi fa sol sol sol fa la sol la fa mi la sol.

used to determine the pitch of the rates of beating of mean tone temperament and the system of equal harmony he used "D-sol-re," while to describe the note above middle C he used "d-la-sol-re."

To continue this story in detail till we reach the full conception of classical tonality towards the end of the eighteenth century is unnecessary, for the curious will find it set out, in the clearest possible manner, in the article "Harmony," by the late Sir Donald Tovey, in the fourteenth edition of the 'Encyclopædia Britannica.' The use of tonality as a basis of the architecture of an extended musical composition in the classical period is there made evident.

This excursion into the early history of English music is to be excused as necessary to an understanding of the problem of the keyboard instrument as a whole, a problem which is at least 400 years old, and one that brings us back to the familiar domain of physics. For the problem of tuning is essentially a problem of physical acoustics, since it depends on beats. The outstanding feature of temperament, as a method of tuning a keyboard instrument, is the rigid intonation it tries to impose on the flexible scale of music. On the other hand, the problem of the musical scale, in its scientific aspect, is a problem in the physiological and psychological sections of the theory of hearing; and it was Helmholtz's great contribution to musical acoustics that he insisted on this truth. The great advance made by his contribution is the more evident when seen against the theoretical background of his day, particularly as it appeared in this country. The errors of nineteenth-century "arm-chair" theorists made it difficult for his contemporaries to appreciate his contribution to the full; but to understand their errors is to excuse them. For this earlier music was practically a sealed book to them. The outlook of the nineteenth-century theorist was limited to music beginning with Handel and ending with Beethoven.

Nor was this all. In so far as they were at all aware of earlier musical history the nineteenth-century theorists saw in it only a primitive stage in the progressive development of music which reached its goal in the music of the classical period. They were unaware that the achievements of the seventeenth-century pre-classical period had really involved the loss of much of the achievement of the sixteenth-century polyphonic period, with its subtleties of rhythm and its instinct for euphony. Assuming that scales came before music, and that temperaments were musical scales, they were deeply concerned about the harmful effect which the adoption of equal temperament, for the tuning of keyboard instruments, might have on the art of music. Their ideas were carried to a logical conclusion by their distinguished contemporary, A. J. Ellis, in the explanations of Helmholtz's meaning which he added to his translation of 'Tonempfindungen' as 'Sensations of Tone.' This logical conclusion was, in fact, a *reductio ad absurdum* of contemporary theory. His narrow and rigid conception of tonic relationships was the basis of the fantastic system of "duodenes" which he devised for the guidance of composers. This system was the kernel of his theories. Had Ellis's musical premises been sound, his conclusions would have been a useful contribution to musical acoustics. He must not be blamed because his musical premises were wrong, for they were current dogma

in his day. The particular misconception with which I am concerned, and it is fundamental, was the unqualified use of the term "just intonation" for the cycle with which we began. The use of the term "just temperament" for this cycle by some modern writers is an attempt to correct this misconception, which is a distortion of the meaning of "just intonation" as used by English musical scholars from Thomas Morley to Sir Donald Tovey, and therefore a source of endless confusion in musical writings to this day. How completely this misconception is set right by numberless passages in Palestrina's compositions was the theme of the article "Just Intonation Misconceived," to which I made reference in the first paper of this series*. As that article was published in the issue of 'Music and Letters' for July, 1943, there is no occasion to repeat here the evidence from music and musical writings it contained. But as the conclusions there reached, concerned as they were with Ellis's translation of Helmholtz's 'Tonempfindungen,' may have some interest for men of science as well as for musicians, I reproduce them in the following paragraphs and table by the courtesy of the Editor of 'Music and Letters.'

It is because Ellis secured some permanent record for his theories by setting out his conclusions in a series of interpolations, footnotes and additions to his translation of 'Tonempfindungen' that they need to be corrected to-day. The serious student of Helmholtz who reads 'Sensations of Tone' can avoid them, for they are all carefully distinguished from the authentic text by square brackets—a matter in which Ellis was very punctilious. What Ellis unfortunately failed to realize was that the use of his own conception of "just intonation" in the translation of a number of passages in Chapter XVI. obscured Helmholtz's meaning. This chapter, in 'Sensations of Tone,' has led many musicians to dismiss, on a first reading, the conclusions of Helmholtz's experimental work, to their own loss. Helmholtz was too great a man of science, and too good a musician, to follow in the footsteps of nineteenth-century "theoreticians."† Indeed, he describes them as theorizing friends of music who feast on arithmetical mysticism.

To enable readers to judge for themselves, the passages in question have been set out below in tabular form. In the left-hand column is the original German of the third edition of 'Tonempfindungen.' In the right-hand column is Ellis's translation of this edition which he used for the first edition of 'Sensations of Tone.' Page references are given for both, and significant words are shown in italics. In the middle column is a suggested English translation, in language corresponding as closely as may be to the original German so as to avoid giving it an unintentional colour. For this translation I am indebted to my colleague Miss Dorothy Butterworth. Her collaboration was particularly valuable,

* Phil. Mag. xxxiv. (7) pp. 472 and 478.

† "We must distinguish carefully between composers and theoreticians [*musicalischen Theoretikern*]."—Hermann von Helmholtz,

A.—References to the tuning of keyed instruments.

page	'Tonempfindungen'; 3rd edition	Suggested English translation	page	'Sensations of Tone'; 1st edition
492	bei <i>reiner Stimmung</i> der <i>reinen Stimmung</i>	¹ (1st time) <i>tuning</i> with <i>true</i> intervals	488	<i>perfect intonation</i>
500	die Vergleichung der Wirkungen dieser Stimmung und der <i>reinen</i> [Stimmung] . . .	the comparison of the effects of this tuning with those of <i>true tuning</i> . . .	497	to compare the effect of this tuning with <i>just intonation</i> . . .
500	Was nun die musikalischen Wirkungen der <i>reinen Stimmung</i> betrifft, so ist der Unterschied zwischen dieser und der gleichschwebenden oder der griechischen <i>Stimmung</i> nach <i>reinen</i> Quinten doch sehr bemerklich. Die <i>reinen</i> Accorde . . .	³ As for the musical effect of <i>true tuning</i> the difference between this and the equally tem- pered [tuning], or the Greek <i>tuning</i> with <i>perfect fifths</i> , is very noticeable. The <i>true</i> chords . . .	497	As regards musical effect, the difference between the <i>just</i> and the equally-tempered or the <i>just</i> and the Pythagorean <i>intonations</i> are [sic] very remarkable. The <i>justly-intoned</i> chords . . .
500	Septimenaccorde in <i>reiner Stimmung</i> ausge- führt . . .	⁴ Chords of the dominant seventh <i>tuned</i> <i>true</i> . . .	497	Chords of the dominant seventh in <i>just</i> <i>intonation</i> . . .
498	Aber wenn man sich durch längere Reihen <i>rein gestimmter</i> Accorde bewegt . . .	⁵ But if we proceed through a long series of <i>truly tuned</i> chords . . .	495	But when long series of <i>justly-intoned</i> chords have been employed . . .
498	<i>rein gestimmte</i> Dur- und Mollaccorde . . .	⁶ major and minor chords <i>truly tuned</i> . . .	495	<i>justly intoned</i> major and minor chords . . .
506	von meinem <i>rein gestimmten</i> Harmonium . . .	⁷ from my <i>truly tuned</i> harmonium . . .	503	from my <i>justly-intoned</i> harmonium . . .
507	den Unterschied zwischen <i>rein gestimmten</i> und temperirten Accorden . . .	⁸ the difference between <i>truly tuned</i> and tem- pered chords . . .	504	the difference between <i>justly-intoned</i> and tem- pered chords . . .
512	an dem <i>rein gestimmten</i> Harmonium . . .	⁹ on the <i>truly tuned</i> harmonium . . .	510	on my <i>justly intoned</i> harmonium
628	Plan für <i>rein gestimmte</i> Instrumente . . .	¹⁰ Plan for <i>truly tuned</i> instruments . . .	633	Plan for <i>justly intoned</i> instruments . . .
628	dass man mit einem Manuale in allen Tonarten <i>rein</i> spielen kann . . .	¹¹ that, with one manual, one can play in <i>perfect tune</i> in all keys . . .	633	as to play in <i>just intonation</i> in all keys . . .
492	von denen der <i>reinen Temperatur</i> . . .	¹² from those of <i>true tuning</i> . . .	488	from those of <i>just intonation</i> . . .

A.—References to the tuning of keyed instruments.

‘Tonempfindungen’; 3rd edition	Suggested English translation	‘Sensations of Tone’; 1st edition
page		page
497 Also haben wir auf dem Instrumente vollständig rein die Molltonarten:	13 On this instrument, therefore, we have minor keys which are absolutely true, viz.:	494 Hence the instrument furnishes the following perfectly just minor scales:
494 ein Instrument . . . welches im Stande ist, durch alle Tonarten in reinen Intervallen moduliren zu können.	14 an instrument . . . which is capable of modulation in true intervals through all the keys.	490 an instrument . . . by which I could modulate by perfect intervals into all keys.
503 und der Contrast der reinen Intervalle ganz fortfällt.	15 and for which the contrast of the true intervals is entirely lacking.	500 and are not contrasted with others in perfect intonation.
504 Ausser . . . dem Mangel eines Vergleiches mit reinen Intervallen . . .	16 Besides . . . the lack of a comparison with more true intervals . . .	501 the absence of any power to compare its effects with those of just intonation . . .
506 reine Quinten und Terzen	17 true fifths and thirds	503 just fifths and thirds
494 Tönen, welche wirklich die theoretisch geforderten natürlichen Intervalle mit einander bilden . . .	18 Tones which really form with one another the natural intervals required by theory . . .	490 Tones which really form with each other the natural intervals required by theory . . .
500 die Differenz zwischen natürlichen und temperirten Accorden . . .	19 the difference between natural and tempered chords . . .	497 The difference between natural and tempered intonation . . .
501 der Unterschied zwischen dem temperirten und dem natürlichen System . . .	20 the difference between the tempered and the natural system . . .	498 the difference between tempered and just intonation . . .
496 und man kann sie alle rein in der natürlichen Tonleiter ausführen:	21 and they can all be played true in the natural scale;	494 they can all be played in perfect purity in the natural intonation;
501 Darüber kann keine Frage sein, dass das System der temperirten Stimmung durch seine Einfachheit ganz ausserordentliche Vorzüge für die Instrumentalmusik hat, dass jedes andere System einen ausserordentlich viel complicirteren Mechanismus der Instrumente bedingen würde . . .	22 There can be no question that because of its simplicity the system of tempered tuning is extremely advantageous for instrumental music, that any other system would require an extraordinarily greater complexity in the mechanism of the instruments . . .	498 There can be no question that the simplicity of tempered intonation is extremely advantageous for instrumental music, that any other intonation requires an extraordinarily greater complication in the mechanism of the instrument . . .

B—References to the intonation of voices and strings

	‘ Tonemfindungen ’ ; 3rd edition	
page		
509	Dass solche Spieler . . . nach natürlichen Intervallen spielen, . . .	
509	die Tonleitern nach natürlichen Intervallen zu zu spielen . . .	
509	die Unterschiede in der natürlichen Stimmung gleich benannter Noten . . .	
510	die richtige natürliche Stimmung	
513	Sobald die Violinisten ihre Tonleitern nach richtiger Stimmung der jedesmaligen Leiter zu spielen sich entschliessen, . . . werden auch die übrigen Orchesterinstrumente so viel nachgeben können, dass sie sich der richtigen Stimmung der Violinen anschliessen. Ueberdies haben unter diesen die Hörner und Trompeten schon die natürliche Stimmung.	
511	Wo sollen aber auch unsere Sänger lernen rein zu singen . . .	
511	vollkommen rein klingende Quartetts singen.	
511	dass man Quartetts öfter vollkommen rein von jungen Männern hört, welche . . .	
511	Reinheit des Gesanges . . .	
511	ein rein ausgeführter Gesang . . .	
510	auf den reinsten Wohlklang der Consonanzen . . .	[of It

B—References to the intonation of voices and strings

Suggested English translation		'Sensations of Tone'; 1st edition
23	That such performers . . . play by <i>natural intervals</i> , . . .	506 That performers . . . play in <i>just intonation</i> , . . .
24	to play the scales by <i>natural intervals</i> . . .	507 to play the scales by <i>natural intervals</i> . . .
25	the differences between notes of the same name in <i>natural intonation</i> . . .	507 the differences between notes of the same name in <i>just intonation</i> . . .
26	the correct <i>natural intonation</i> . . .	508 the correct <i>natural intonation</i> . . .
27	As soon as the violinists decide to play their scales in the <i>correct intonation</i> of the actual scale . . . the remaining orchestral instruments will be able to adapt themselves to the <i>more accurate intonation</i> of the violins. Moreover, amongst these instruments the horns and trumpets already have the <i>natural intonation</i> .	511 As soon as violinists have resolved to play every scale in <i>just intonation</i> . . . the other orchestral instruments will have to suit themselves to the correct <i>intonation</i> of the violins. Horns and trumpets have already <i>naturally just intonation</i> .
28	But where are our singers to learn to sing <i>true</i> . . . ?	508 But where are our singers to learn <i>just intonation</i> . . . ?
29	singing quartets with <i>perfectly true intonation</i> .	509 singing quartets in <i>perfectly just intonation</i> .
30	that we more often hear quartets sung <i>perfectly true</i> [or <i>with perfect intonation</i>] by young men who . . .	509 that quartets are more frequently heard with <i>just intonation</i> when sung by young men who . . .
31	<i>Perfect intonation</i> in singing . . .	509 <i>correct intonation</i> in singing . . .
32	a song that is sung <i>true</i> [or <i>with perfect intonation</i>] . . .	509 a song when sung in <i>correct intonation</i> . . .
33	in church music of the fifteenth and sixteenth centuries] for the <i>most perfect euphony</i> of the consonances . . .	508 for <i>most perfect intonation</i> of the chords . . .

for she not only brought to the task an intimate knowledge of German and considerable experience in translating scientific and technical papers written in German, but she was wholly unacquainted with the pseudo-scientific speculations of the nineteenth-century "theoreticians," and was therefore unable—for the best of reasons—to read any trace of them into her translation.

Some comments will be added later; but, by way of introduction, three points need mention. First, the quotations have been re-arranged in two groups, so as to bring together, first those describing Helmholtz's experiments with his specially tuned Harmonium* and then those which deal with the free intonation of voices and strings. Secondly, while Miss Butterworth used "tuning" throughout in her first draft for *Stimmung*, we agreed to follow Tertis's lead† and use "perfect intonation" for strings played in perfect tune. On the other hand, for keyed instruments we avoided "perfect tuning" because it might possibly suggest a piano perfectly tuned in equal temperament. We adopted "true tuning" instead, and to define it clearly we used "tuning with true intervals" the first time. Finally, the reader would do well to refer at the outset to example 18, which explains Helmholtz's definition of "natural intervals" as intervals selected by his theory of consonance which was based on the properties of the ear and the effects of beating.

Comparison between the two translations is interesting; but the only valid test of the suggested translations is to enter them in a copy of the first edition of 'Sensations of Tone' and then judge whether they correctly disclose Helmholtz's own thoughts. For this purpose they must be read in their context, and the chapter itself must be read in its scientific setting. The test is very convincing to me. Why Ellis's outlook differed from Helmholtz's emerges clearly from one of his footnotes‡ in which he wrote:

It is precisely here . . . that those who study just intonation on paper or on instruments without fixed tones (as the voice and the violin) differ from those who study it on instruments with fixed tones. The former find no difficulty in the leap [by which Ellis referred to an interval that demanded an enharmonic change, as Helmholtz had explained in the text]. The latter find it unnatural and impossible. *And so it appears to me.*

The italics are the present writer's, and they sum up the argument of this paper in six words.

* *

The quotations made in this paper from the first edition of 'Sensations of Tone,' for comparison with the original German, are here reproduced

* In example 14 Helmholtz describes, at the outset, the objective of this instrument. Observe the use of the word interval.

† In his 'Beauty of Tone in String Playing.'

‡ Translator's footnote to 'Sensations of Tone,' 1st ed., p. 512.

with the ready assent of Messrs. Longmans, Green & Co., the publishers of 'Sensations of Tone,' for which the writer desires to record his cordial thanks.

NOTE.—I hope that musical readers will form their own judgments from the evidence, in music and musical writings, discussed in my article "Just Intonation Misconceived"; and as there are available very few extra copies of 'Music and Letters,' owing to paper shortage, the Editor and the Printer were good enough to arrange for me to have off-prints, and some of these have been placed at the disposal of Messrs. Taylor and Francis, the publishers of this Journal, who have kindly agreed to send copies to applicants, at a charge of 3*d.* post free.

Gerrard's Cross,
Bucks.

LXXX. *The Visual Spectrophotometry of Glasses, with Special Reference to Low Absorptive Glasses* *.

By T. H. WANG, M.Sc., D.I.C. †

[Received March 24, 1943.]

ABSTRACT.

The types of spectrophotometer designed primarily for the measurement of the light absorption of homogeneous glass are generally not suitable, without modification, for the measurement of low absorption, especially in the case of glasses in which cords and strain are markedly present. A simple analysis shows that when using a small-sized source of illumination, such as the Pointolite lamp, the thickness of the specimen may have to be limited to 0.7 cm. On the other hand, if a broad uniform diffusing source is available, specimens up to 13 cm. thick can be studied without causing serious error. By this means, specimens in the form of rods and sheets can be studied by allowing the illuminating rays to be reflected internally. An apparatus constructed on this principle is described and typical results are recorded. Its application to problems of low absorption, such as occur in the processes of decolorizing glasses, is discussed. The possibility of using comparatively thick glass specimens with Hardy's recording spectrophotometer is pointed out.

I. INTRODUCTION.

In the spectrophotometry of transmission or absorption of transparent and non-diffusing materials, such as glass, one of the main factors influencing the reliability of the measurements is the possible presence

* A Paper read at the Sheffield Meeting of the Society of Glass Technology, Nov. 18th, 1942, and published also in the Journal of that Society.

† Communicated by Prof. W. E. S. Turner, F.R.S.

of optical imperfections in the specimens tested. The conditions generally involved require that the specimen should be homogeneous in respect to both absorption and refraction, and the surfaces should be plane and parallel; otherwise the light rays passing through the specimen may be deviated from their normal paths in an unknown manner, resulting not only in change in the length of the optical paths of the rays, but, what is more serious, the photometric conditions may also be upset.

In the case of glass, particular emphasis should be laid on its homogeneity and therefore uniformity of refractive index. For it is a comparatively easy matter to obtain optically worked surfaces on glass; it is also not difficult to obtain reasonable homogeneity in absorption to comply with the standard of accuracy obtainable in this method. But glass, with few exceptions, is far from optically perfect, mainly due to the almost unavoidable presence of striæ and cords. This is particularly true in laboratory practice, when glasses are melted on a small scale. In certain instruments designed primarily for perfect specimens the permissible amount of imperfections involved may be very small. Often the requirement in homogeneity in refractive index is far more stringent than the required homogeneity in absorption. It is frequently impossible to achieve this requirement in ordinary practice. As a compromise, the specimen to be studied is usually kept thin—say, a few mm. thick—so that any undue deviations of the illuminating rays through the specimen may be reduced. But it may well be the case that specimens at least 10 cm. thick are needed in order to obtain a measurable amount of absorption.

It has, however, long been realised* that by the use of a diffusing source of illumination of uniform brightness and of sufficient extent, the absorption of thick and imperfect specimens as well as lenses and optical instruments may be measured accurately. Such a method should also be possible to solve the main trouble in the case of glass. In the past year, an instrument has been set up in the Department of Glass Technology based on this principle. In view of its suitability for studying problems such as the decolorization of glasses, a brief description may be justified, although only a few minor points in the design are new.

Acknowledgment should be made of an optical arrangement devised by K. Rosenhauer and F. Weidert† suitable for the spectro-photometrical measurement of glasses rich in striæ. It was, however, for thin specimens only.

II. THEORETICAL CONSIDERATIONS.

Let us first refer to some of the most common types of spectro-photometric equipment and examine how the optical inhomogeneity present in the test specimens may influence the optical performance.

* K. S. Gibson, *J. Opt. Soc. Amer. and Rev. Sci. Instr.* x. p. 1698 (1925),

† *Glastech. Ber.* xvi. pp. 51-57 (1938).

Example 1.

Fig. 1 is an optical equivalent diagram of a very common type of spectrophotometric illumination. For the sake of simplicity only the test beam is represented. The photometer part is also not included, and only the collimator of the spectrometer is shown. The source consists of either a Pointolite or of a ribbon filament lamp. For the sake of easy illustration, we further assume that the image of the source is focused on the collimator slit. The specimen is placed in between the two lenses, so arranged that the light between them is parallel. The viewing field at the other end of the spectrometer may be either of juxtaposed-spectra or the Maxwellian-view type. In the former instance the viewing field is the image of the slit; in the latter case the slit corresponds to the viewing pupil.

For an optically perfect specimen, moving the specimen in or out of the beam will not alter any of the ray-paths. On the contrary, any presence of refractive index fluctuation at a certain point P (fig. 1) inside the specimen can be regarded as equivalent to the action of a small prism. Rays passing through it will therefore be diverted from their

Fig. 1.

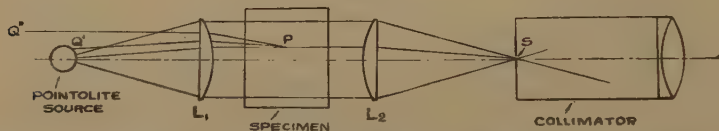


Diagram of one type of optical arrangement, showing the influence of imperfections present in the specimen on the photometric performance. If the ray SPQ does not meet the source, no light would reach S from the point P .

normal paths. Consider the problem photometrically. Let a hypothetical ray be traced backwards from the slit, S , passing through the point P . If the prismatic effect so encountered is small, this ray may still meet somewhere on the source as illustrated by the line PQ' . If the deviation is considerable, this ray will, as represented by the line PQ'' , not meet the source at all. In the latter circumstance, rays from any part of the source passing through P will not enter the slit S . Consequently, in the juxtaposed spectra type of field the brightness of the test beam will be correspondingly lowered. In the Maxwellian-view type of field it will appear non-uniform. Assume that the source is uniformly bright all over its luminous surface. So long as all the possible "reverse rays" can strike upon the source, the photometrical condition is unaltered.

Based on this criterion, we can form a rough estimate of the stringent optical requirement as follows: In general, the radius of the incandescent sphere of the Pointolite lamp or the half width of the ribbon filament is of the order of 1.5 mm. The focal length of the lens L_1 can be taken as 10 cm. The separation between the lenses L_1 and L_2 is in general

not less than 10 cm. Let us assume, for the sake of simplicity, that the overall prismatic effect for a particular ray be originated from one single spot, say P, midway between L_1 and L_2 . Accordingly, the distance between P and the source would be 15 cm. Remember also that the slit has length and width. Thus, to safeguard consistent photometry, the maximum angular diversion for any useful ray should not exceed $\frac{0.15-0.05}{15} = 0.007$ radians or 0.4 degrees, the quantity 0.05 being an estimated reduction due to slit width.

For a glass carefully prepared in the laboratory, a fluctuation in refractive index of the order 0.001 is quite common. When it arises from striæ and cords, such a change may well occur within a distance of 1 mm., corresponding to the existence of a gradient in refractive index of magnitude 0.01 per cm. Denote the thickness of the plane parallel specimen by x . In the most unfavourable case a ray may encounter this gradient throughout the whole thickness and its overall deviation will be $0.01x$ radians. Take x as the maximum allowable thickness, and equate this to the maximum permissible deviation without upsetting

Fig. 2.



Diagram of another example.

The specimen is placed near to the slit. Cone A illustrates the ray paths when the specimen is perfect. Cone B is deviated by any prismatic effect of the specimen, the photometric condition being still promising. Cone C has been deviated too much, the photometric condition becoming upset.

the photometrical condition. We obtain $0.007 = 0.01x$ or $x = 0.7$ cm. Actually, the refractive index gradient may be much greater than we have assumed. On the other hand, the most unfavourable occurrence is also scarce, provided that due precaution has been taken in selecting the specimens. The value 0.7 cm. accords well with our usual experience.

Example 2.

In this instance, as illustrated in fig. 2, the specimen is placed close to the slit or its image. Again the image of the Pointolite lamp source is focused on to the slit by the lens L_1 . Virtually, this lens forms the apparent source of a size equal to its aperture. A field lens L_2 may be placed in close contact with the slit to focus the image of the apparent source on the objective of the collimator. If we restrict ourselves to the juxtaposed-spectra type of field, the objective of the collimator is then also the pupil of the spectrometer. Any ray capable of reaching this objective should therefore enter also the pupil of the observer's eye or any equivalent receptor.

Case 1.

When the aperture of the lens L_1 is small, the cone of rays entering the slit is narrow, and falls only on the central part of the collimator objective. Consider the problem from the point of optical inhomogeneity alone. Any refractive index gradient present in the specimen at the part under consideration will cause the cone of rays to deviate as a whole from its normal position. If there is no error introduced in absorption due to the deviated rays passing different portions of the optical system (for example, passing nearer to the base of the refracting prism of the spectrometer and entering the eye through the periphery of the pupil), we may say that the photometrical condition remains true so long as no part of the cone of rays becomes inaccessible to the pupil. Assuming the focal ratio of the collimator as $f/10$, one can readily see that the permissible amount of imperfection in the specimen can be much more generous in comparison with the first instance (fig. 1), provided that the aperture ratio of the lens L_1 is kept low.

Case 2.

An alternative method is to increase the aperture of the lens L_1 sufficiently so that the entrance pupil can be fully filled with light not only when a good specimen is used, but also remains so when the specimen is not homogeneous. This is essentially the device used by K. Resenhauer and F. Weidert*. In this arrangement, unlike that in Case 1, the rays inside the instrument are not deviated when the specimen is inserted in the test beam. Thus there will be no variation in absorption due to changed condition of rays, as can happen to Case 1.

In both cases, referring to fig. 2, thick specimens cannot be used with advantage; for, not only would the apparent size of the source be changed, but also, since different parts of the specimen will be at different distances from the slit, the limitations described in Example 1 will arise.

The author gives the above two examples mainly to illustrate how the optical set-up itself may set a limit to the size of the specimen which can be employed. Any other arrangement can be dealt with in a similar way, and will in general not involve any new principles.

In the examples given, the situation will, however, be entirely changed if the source can be extended and be uniformly bright not only in area, but also in angular subtense. Thus, all useful rays when traced backwards will meet the source at the same brightness. For instance, the lens L_1 in fig. 2 can be discarded altogether. Since the arrangement would then become insensitive to refractive index inhomogeneity or surface imperfections, the thickness of the specimen, if necessary, could be increased considerably, to that extent at which the optical path through the specimen becomes so distorted that the change in effective length would influence appreciably the absorption measurement.

* *Loc. cit.*

Let us set up a criterion such that in the most unfavourable circumstance the length of path of any useful ray shall not be lengthened by more than 1 per cent. of the specified length. This corresponds to the fact that the rays should not be inclined to the optical axis by more than 7° . Actually only a small part of the useful rays would be altered to such an extent. Depending on the method employed, the result observed can be well made to average out the more pronounced discrepancies. Hence the error introduced due to this factor may be made sensibly smaller, say, within 0.5 per cent.

In the case where the specimen is placed directly in front of the slit or its equivalent position, the rays inside the specimen primarily cover a solid cone. With an aperture ratio of $f/10$ this cone subtends an angle of 2° about the axis (assuming an index of refraction for the specimen of 1.5). There remains thus $7^\circ - 2^\circ = 5^\circ$ to be allocated to the imperfections of the specimen. In the previous paragraphs we have already noted that for an external deviation of 0.4° a thickness of 0.7 cm. is permissible for a glass specimen with perfect optical surfaces. Again, if we attribute the whole remaining allowance to refractive inhomogeneity, the maximum

Fig. 3 A.

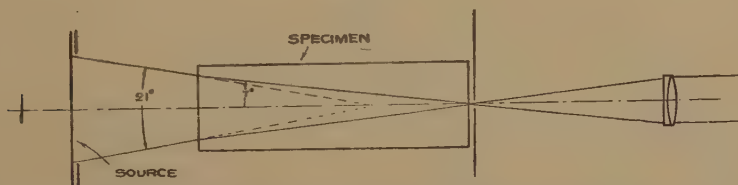


Diagram illustrating the use of a broad uniform diffusing source, allowing the thickness of the specimen to be increased considerably.

length feasible would be $\frac{5 \times 1.5}{0.4} \times 0.7 = 13$ cm. Alternatively, by restricting ourselves to the use of somewhat shorter specimens—say, 10 cm.—the presence of some surface imperfections would become allowable.

With the maximum inclination of 7° for the rays inside the specimen, all the useful rays would be confined within a cone subtending 21° towards the source (fig. 3 A). What we require is therefore a uniformly bright source of sufficient area and perfectly diffusing—*i. e.*, following the cosine law of light distribution, within the normal cone of 21° . The acquirement of such a source is not difficult, but will be even easier to arrange if its size can be further reduced by using the arrangement shown in fig. 3 B. In this case specimens of thin cross-section are used instead of glass blocks. The side walls of the specimen are polished plane parallel or cylindrical, so that rays will be critically reflected inside, and will finally emerge in exactly the same way as before. The great advantage of this method is, however, that glasses can be studied in the form of sheets or rods.

III. THE EXPERIMENTAL JUSTIFICATION FOR USING THICK SPECIMENS, ESPECIALLY RODS.

Particular reference may be made to the use of rods. In systematic measurements on low absorption glasses, glasses are usually prepared from the purest materials. The use of rod specimens greatly reduces the amount of glass needed as compared with glass blocks. This not only saves the labour in preparing them, but also makes the control of other factors much easier. Furthermore, glasses drawn in rod form give naturally fire-polished cylindrical surfaces, often quite uniform and smooth enough to warrant their direct application to such work. The remaining labour in preparation of the specimen consists only in shaping and polishing the ends. As pointed out, by using specimens not more than 10 cm. in length the presence of some surface imperfections is permissible. Adhering to the practice of using rods, part of this allowable imperfection should be attributed to the possible non-straightness and non-uniformity of the drawn surface. Moreover, it is a familiar fact that, with due care, glass rods of 4 or 5 mm. diameter can be cut square with a reasonably good end-surface by a single hand operation. If such a surface is good enough to justify its direct use for the present work,

Fig. 3 B.



Diagram showing the reduction in size of the broad source by the use of rod or sheet specimens.

the labour and time required in preparing the specimen are reduced very greatly.

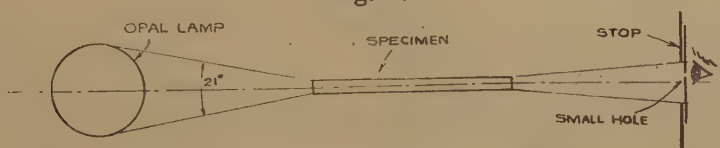
To justify the procedure referred to, a simple test has been carried out as follows :

A glass rod of 4 mm. diameter was prepared from a laboratory melt (usually the uniformity in cross-section of such rods is not comparable with the commercial ones). It was first inspected for freedom from seeds and inclusions, and then cut with care into 10 cm. lengths. After this the end-surfaces were further inspected to ensure that they were well squared, had no irregular chippings, no pronounced filing marks, etc. This rod was then placed in position as illustrated in fig. 4. An opal lamp was placed in alignment with the rod such that all the light reached the front end surface within a cone of 21° . The other end of the rod was then viewed at a certain distance through a small hole which could be shifted to any position on the stop. The size of the stop corresponded to the actual aperture ratio at this end when actual measurement was taking place. If the conditions cited were satisfied,

one would in all cases see the end surfaces uniformly filled with light. Otherwise, some part at the end would appear dark, corresponding to the absence of those rays which would have been deviated by an amount greater than could normally be allowed. The result indicated that the requirement is easy to achieve. According to the author's experience, only a very small fraction of the specimens examined has been found unusable in this procedure. It should also be stated that the occasional appearance of small thin streaks as thus observed would do little harm so long as a slight opening of the illuminating cone can make them disappear. In the present instance this can be done by shifting the lamp nearer to the rod. Since the actual source to be used should be wider in extent, the error so introduced will be inappreciable.

The above simple test is regarded as necessary before the specimen is taken for measurement.

Fig. 4.



Method of inspecting a specimen for its optical performance.

IV. THE DESCRIPTION OF THE APPARATUS.

The apparatus is intended for visual spectrophotometry and is schematically represented in figs. 5A and 5B. What is novel in it consists mainly of the addition of the illuminating device.

A fundamental requirement is that the source should be as bright as possible, so that a narrow slit can be used for the spectrometer. The source, which should be uniform and diffusing, consists of a 150-watt coiled coil-filament opal lamp A. The diameter of the bulb is 10 cm. On account of the fact that only a very small portion of its surface is in actual use, the required uniformity is not difficult to secure, first, by a pre-selection of the lamp to be used, and, secondly, by a choice of the best part of the surface. The uniformity and brightness are further enhanced by enclosing this lamp in a diffusing sphere B of 20 cm. diameter, the inside surface of which is coated with magnesium carbonate. The essential part of the opal surface is ground matt, so that no image of the window of the diffusing sphere is formed on the source.

Two beams of light are derived, one test beam and one comparison beam. Each passes through a separate hole 3 mm. diameter on the screen E. These two holes act as virtual sources and are focused on to the slit of the spectrometer by a pair of de-centred lenses G_1 and G_2 . In between these and the slit is the polarizing photometer H, which also brings the two beams juxtaposed on the slit. A pair of stops placed immediately behind these de-centred lenses forms the entrance pupil of the spectrometer.

Fig. 5 A.

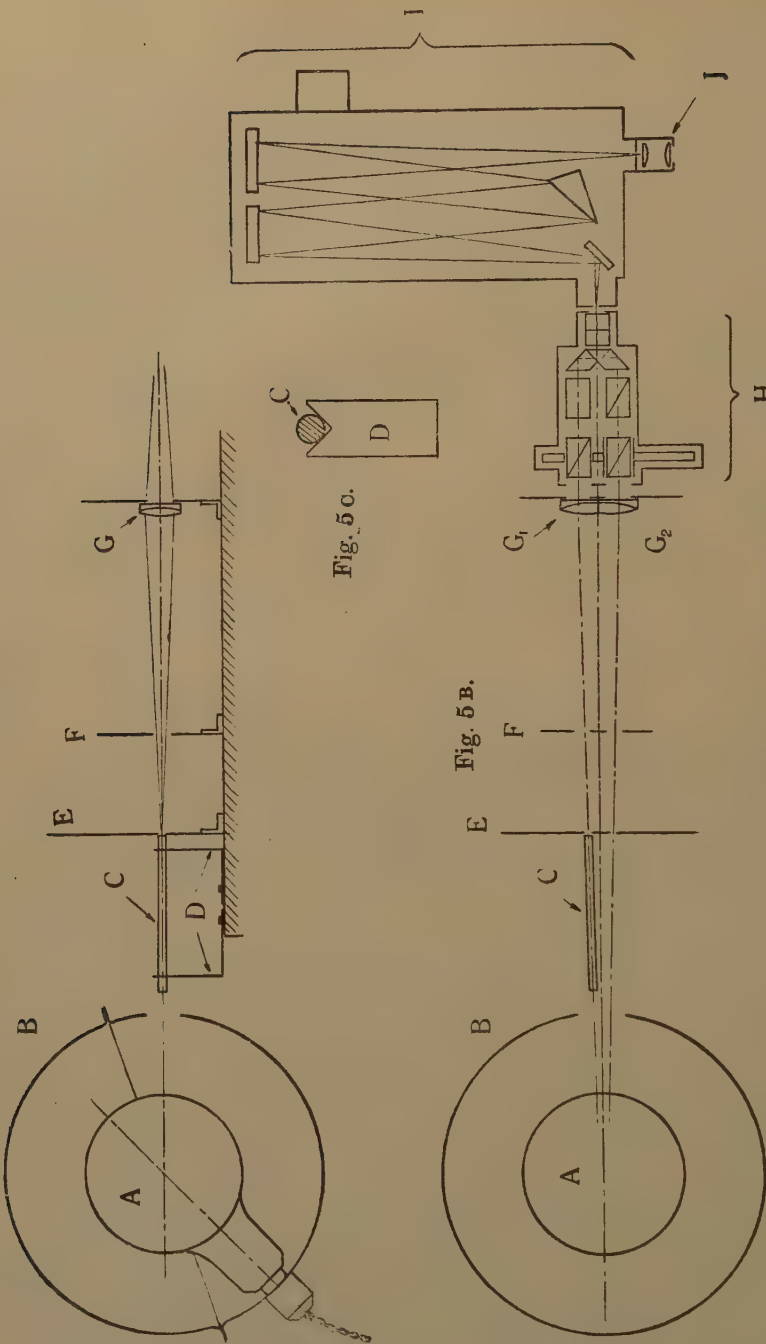


Fig. 5 C.

Fig. 5 B.

Schematic diagrams of the apparatus.
(For lettering, see p. 693.)

beam and can be replaced in position by means of locating studs. Several racks are provided to take specimens of different lengths. The source can be moved forward or backward to suit the length of the specimen. In the closest position the window of the source can be as near as 5 cm. from the screen E. As a suitable position of the source the sphere-window is usually positioned 1 cm. distant from the rear end of the specimen.

The stop F is to limit both the test beam and the comparison beam, so that no superfluous light from one can reach the aperture of the other. Any stray light unsymmetrically disposed in the two beams is thus eliminated.

In the present instance the polarizing photometer and the constant deviation spectrometer were manufactured by Bellingham and Stanley, Ltd. The intensity of the comparison beam is rendered variable by rotating the polarizer, and the transmission value follows theoretically the \cos^2 relation. In the spectrometer, aluminized concave spherical mirrors are incorporated as collimator and telescope. The photometric field is the juxtaposed spectra type. In view of the fact that the brightness of the present source is much weaker than that of an ordinary Pointolite lamp, the entrance pupil is completely filled with light. Also, for ordinary measurements, where no sharp change in absorption with respect to change in wave-length takes place, the collimator slit is opened comparatively widely corresponding to a wave-length range of 20, 60, and 90 Å. at wave-length 4500 Å., 5893 Å., and 6500 Å., respectively. The viewing field is opened to the same extent. By using an eye-piece of slightly greater magnifying power than the one originally provided on this instrument, the photometrical field subtends $1\frac{1}{2}^\circ$ horizontal and $2\frac{1}{4}^\circ$ vertical. The major diameter of the exit pupil (slightly elliptical) is $2\frac{1}{2}$ mm. Since the effective size of the exit pupil is on the high side, in order to fix readily the eye-pupil on to it, the dental-impression method for fixing the observer's head is incorporated*. The author found that with this it is much easier to obtain more consistent results; for, under such conditions, the two beams always completely enter the eye.

* This device has been advantageously used by W. D. Wright on his colorimeter (*J. Sci. Instr.*, xvi. p. 10 (1939)).

Fig. 5 (lettering).

Fig. 5 A, side view. Fig. 5 B, plan view. Fig. 5 C, method of supporting the specimen. In these diagrams:

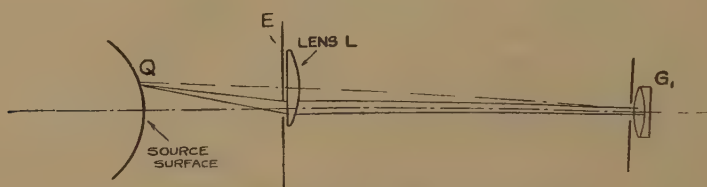
- | | |
|--------------------------------------------|---------------------------------------------------------------------|
| A. 150 Watts opal lamp. | G. De-centred lens. |
| B. Diffusing sphere. | H. Bellingham and Stanley polarising photometer. |
| C. Rod specimen being tested. | I. Bellingham and Stanley wave-length spectrometer. |
| D. Specimen rack. | J. Eye-piece of slightly bigger magnifying power than the original. |
| E. Light screen. | |
| F. Stop for eliminating undue stray light. | |

V. THE TESTING OF THE APPARATUS.

Apart from the usual calibrations necessary, such as the wave-length calibration, the calibration of the photometer scale, etc., the performance of the present apparatus has been checked mainly from the viewpoint of the fundamental requirement—namely, provision of a broad uniform source of illumination. This has been tested in the following manner:—

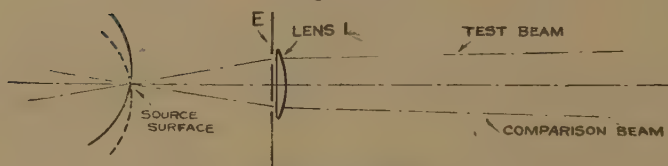
As shown in fig. 6 A, the surface of the source was placed at the focus of a lens L, which covered the aperture of the test beam on the screen E. This beam was further narrowed by stopping down the aperture of the focusing lens G_1 . Thus, all the rays of the narrowed beam originated from a particular spot Q on the surface of the source. This lens L, having a size slightly greater than that of the useful area of the source, can be shifted laterally with respect to the test beam. In consequence,

Fig. 6 A.



Testing brightness uniformity in extent.

Fig. 6 B.



Testing diffuseness.

Diagrams showing the methods of testing the source.

the position of Q shifts accordingly. By making photometric matchings at various positions of Q, the brightness distribution over the useful area on the source can be surveyed. This was carried out along two perpendicular diameters at three different wave-lengths, 4600 Å., 5500 Å., and 6800 Å. In twenty-five measurements made, the maximum deviation from the average was 0.7 per cent. In the present set-up, however, the area covered on the source is practically the same for both cases according as the specimen is in or out of the beam. Consequently, the error introduced in a measurement will be much less.

There may be another source of error due to the possible departure from perfect diffuseness of the source within the cone of emergence under consideration. To test this, a lens (fig. 6 B) was placed symmetrically on the screen E, covering both apertures of the test beam and the comparison beam. The source surface was again placed at the focus of

this lens. The two beams thus originated from the same spot, but were inclined at an angle of 10° to one another. The source was first turned with its surface normal to the comparison beam. A photometric setting was taken. Next, the surface of the source was turned to the test beam and another setting made. No observable difference between the two could be found.

In the present installation the transmittance of the comparison beam is varied by rotating the polarizer. Consequently, the possible presence of partial polarization of the source will influence the photometer scale. Calibration of the instrument indicated that the error introduced by this effect is quite small.

VI. OTHER PRECAUTIONS AND SOME TYPICAL RESULTS.

As already mentioned, it is necessary to follow the procedure described in section III. in order to ensure the optical performance of the specimen. In addition, it is essential that the side surface is clean enough so that all the rays will be critically reflected and not scattered. To have very clean surfaces is an obvious necessity, even in ordinary absorption photometry. This is particularly important in the case of long rods where the absorptions studied are intrinsically low. Therefore, the specimen should always be cleaned thoroughly in the optical sense, in particular from grease.

Glasses of special low durability, if studied in the form of rods, should be tested only when the surfaces are freshly prepared. If this is not possible, it is advisable to resort to the use of blocks, the thickness being duly reduced to accommodate the size of the source.

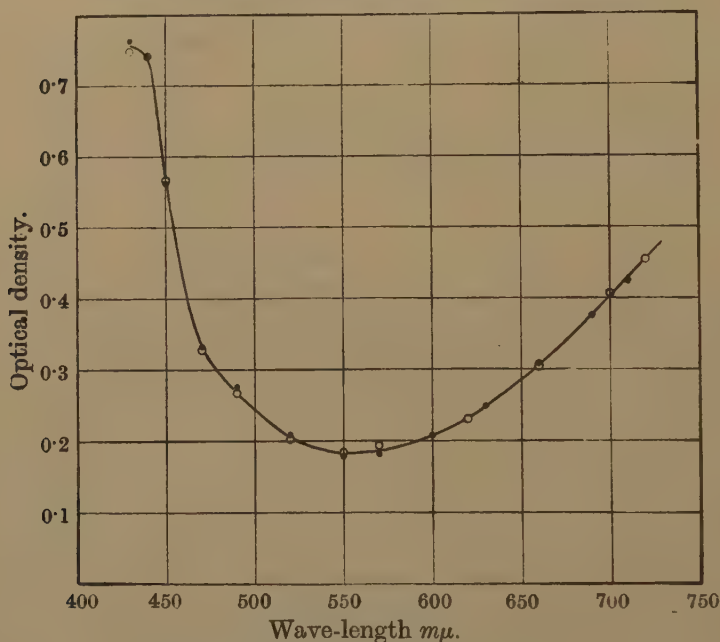
Whenever specimens longer than 10 cm. are used, unless the glass is of special optical quality, only those with straight surfaces and uniform cross-sections should be used. The end-surfaces should always be optically polished, otherwise the required performance might not be readily achieved. Sometimes rods slightly conical in shape can be used, in which case the thicker end should always face the source.

As will be shown by some of the typical results obtained, provided that all the necessary precautions are taken, the relative accuracy of the measurements, either in absorption or transmission, conforms well with the standard which an ordinary spectrophotometer can achieve with a perfect specimen. Therefore, despite the insensitiveness to refractive index inhomogeneity, the accuracy attainable calls for the use of specimens very homogeneous in absorption. This can be obtained, especially in laboratory practice, only by very careful manipulation in preparing the glasses.

The consistence and trustworthiness obtainable in the present instrument are illustrated by the results shown in fig. 7, wherein the spectral absorption of a glass rod of 10 cm. length was obtained in two different ways. The rod was of soda-lime-silica glass containing 0.26 per cent. of iron, determined as Fe_2O_3 , and 2.3 per cent. of arsenic, determined as

As_2O_3 , the glass having been melted in a gas-fired furnace. The absorption of the full-length specimen was determined in the usual manner. The results are indicated as circles on the graph (fig. 7). Next, the same specimen was divided into two halves, the spectral absorption of each half being measured separately. After a deduction of the loss of light due to surface reflexion calculated from the Fresnel fraction $\left(\frac{N-1}{N+1}\right)^2$, the two absorptions so obtained were added together. The results are indicated as black dots on the same graph. In neither cases were the specimens end-polished. The concordance of these two sets of results can be well seen on the graph.

Fig. 7.
Results illustrating the reliability of measurement.



Length of specimen = 10.08 cm.

\circ = Results from full-length specimen.

\bullet = Summation results from the two halves.

Spectral absorption of the glass No. 213 As_2O_3 3 (soda-lime-silica glass containing 0.26 per cent. iron oxide, 2.3 per cent. arsenious oxide).

In fig. 8 are shown the spectral absorption curves (curves 1 and 2) of two soda-lime-silica glasses containing 0.04 and 0.02 per cent. of iron, respectively (determined as Fe_2O_3), the glasses being melted in a platinum crucible for 20 hours in an electric furnace. In both instances the lengths of the rods used were 10 cm. The rods were end-polished.

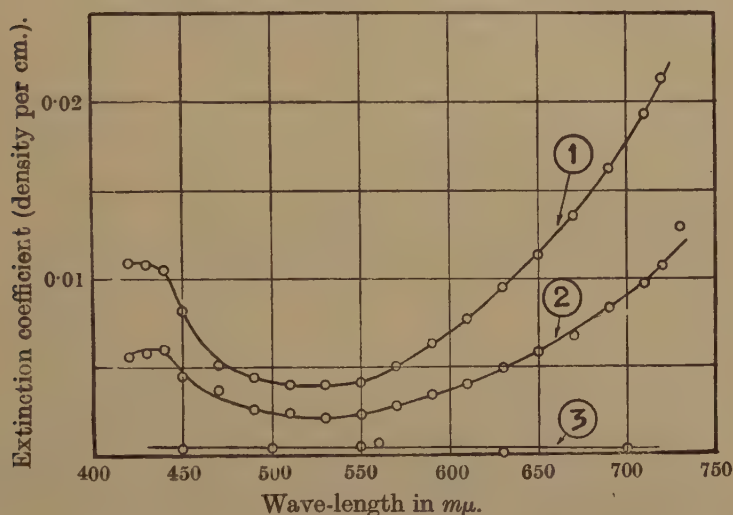
Near the bottom of the same graph are shown the results (curve 3) for a soda-lime-silica glass prepared from pure materials (quenched

cleaned quartz, A.R. sodium carbonate, A.R. precipitated chalk), the mixed batch being sintered and melted in a platinum crucible. The slight absorption indicated is probably due partly to a slight solution of platinum in glass, partly to the presence of traces of impurities, and to a slight degree to small imperfections of the end surface, since the rod was not end-polished.

The satisfactory brightness of the viewing field at the slit-width mentioned (60 A. at 5893 A.) is illustrated by the fact that measurements can be carried out at wave-lengths 0.42μ and 0.73μ at the two ends of the visible spectrum, though needing careful work to accomplish.

Fig. 8.

Some typical results of low absorption glasses.



Curves 1, 2, and 3 are the spectral absorptions of soda-lime-silica glasses prepared from pure materials containing 0.04, 0.02, and no iron oxide, respectively. In all cases the reflection in measurements were deduced from calculations. The small amount of absorption present in curve 3 may be due to various reasons.

VII. CONCLUSIONS.

The arrangement described was intended as a temporary set-up for investigations associated with the decolorizing of glasses. Some improvements could be made if another set were built. For example, the brightness of the source could probably be much increased by the use of a more powerful source. The size of the diffusing sphere might be reduced and its shape changed. In certain respects a diffusing reflecting source may be better than a diffusing transmitting source as used in the present set-up. A lens might be attached to the sphere-window so that the effective area of the source could be further reduced, and, consequently, the apparent uniformity of the source increased.

The problem of the decolorizing of glass has been tackled by different authors in various ways. One method used to a wide extent is to enhance the colouring effect by increasing the concentrations of the colouring agents in question. The spectral absorption can then be measured with comparative ease by using thin specimens*.

Whilst the information so obtained is of interest, there must often be doubt as to how far the results can be representative of the true conditions where the colouring effect is low. The simultaneous presence of both colouring and decolorizing agents presents another difficulty to this method, since the colouring effect depends sensibly on the concentrations of both. Investigators concerned with the practical aspects of decolorizing have, therefore, resorted to visual inspection†. The results had usually been expressed by terms such as "pale blue" "light green," etc., and the conclusions drawn went no further than stating whether the effect was "physical" or "chemical." In some cases‡ the results were expressed in colour coefficients from the observations made on a colorimeter. Such results should, however, be used with care, especially in correlating other physical facts. The statements need to be very exact, otherwise the result only causes confusion; for example, a "blue component" in additive colorimetry and in subtractive colorimetry mean entirely different things. All these facts only show the necessity for an instrument capable of measuring accurately low spectral absorptions.

A number of attempts have been made in the past to provide such an instrument. In 1926, a description of a photo-electric spectral absorption meter for measuring the absorption of optical glasses appeared in the literature§. This instrument was further improved by the incorporation of a monochromator||. Ever since the invention of the barrier type of photo-electric cell, one of its first applications was to measure the spectral absorption of coloured glasses. Attempts were also made to measure the spectral absorption of optical glasses and ordinary "colourless" glasses¶. From the point of convenience, this method is ideal. On the other hand, the reliability of the result measured still depends on a sound optical arrangement. Recently, the development of the recording photo-electric spectrophotometer** in the United States

* For example, W. E. S. Turner and W. Weyl, *J. Soc. Glass Tech.*, Trans. xix. p. 208 (1935); W. Weyl, *Sprechsaal*, lxxi. pp. 7, 91, 104, 117 (1938).

† For example, N. E. Densem and W. E. S. Turner, *J. Soc. Glass Tech.*, Trans. xxii. p. 372 (1938).

‡ For example, E. J. Gooding and J. B. Murgatroyd, *J. Soc. Glass Tech.*, Trans. xix. p. 43 (1935).

§ W. D. Haigh, *Proc. Opt. Conv.*, Part I. p. 327 (1926).

|| W. M. Hampton and W. N. Wheat, *J. Soc. Glass Tech.*, xv. p. 306 (1931), particularly p. 312.

¶ W. Eitel and B. Lange, *Glastech. Ber.* ix. p. 211 (1931).

** A. C. Hardy, *J. Opt. Soc. Amer.* xxviii. p. 360 (1938); J. L. Michaelson, *ibid.* xxviii. p. 365 (1938).

marked a great advance in spectrophotometry in regard to both accuracy and convenience, and several publications describe its application to problems of glass transmission and colour*. According to what the author has learned from various descriptions, in this recording instrument the matching of the comparison beam and the test beam is accomplished in a diffusing sphere. Consequently, in a transmission or absorption measurement, so long as all the useful flux can enter the sphere, the photometric condition is satisfied. In other words, this instrument is "insensitive" to refractive inhomogeneity due essentially to the same reason described in the present paper. Thus, it is quite feasible to use comparatively thick glass specimens on this instrument.

The use of sheet and rods would be advantageous in studying the effect of the solarization of slightly coloured glasses. Recently, there has been evidence that the degree of solarization depends, apart from other factors, on the ultra-violet absorption of the glass. Consequently, to obtain the full effect of solarization, the specimen under exposure should be thin. On the other hand, to observe the effect efficiently, the specimen should be thick†. Both requirements can, however, be satisfied by exposing the thin specimen sidewise and observing it lengthwise.

For some time, in the Department of Glass Technology, researches have been in progress on the influence of temperature on the spectral absorptions of glasses and some of the results have already been published‡. The present apparatus, due to its insensitivity to refractive inhomogeneity and surface conditions, affords a possibility of studying the absorptions above the softening point of the glass, in which region the specimen may become slightly deformed due to softening. In view of the meagreness of physical data in this semi-softening range, this suggestion may be worth trying.

Finally, the author wishes to note that the principle involved in the present apparatus is so simple that most of us have utilized it without noticing it. Thus, when we examine the colour of a glass rod or a slab, we are accustomed to view it against a broad uniform background instead of a small white patch, because only then does the colour appear uniform.

In conclusion, the author is indebted to Professor W. E. S. Turner for his encouragement and interest during the setting up of the present arrangement, and to the Glass Delegacy for the Research Fellowship which enabled this research to be undertaken.

Department of Glass Technology,
The University, Sheffield.

* A. E. Badger and A. C. Ottoson, *J. Amer. Ceram. Soc.* xxv. p. 104 (1942).

† L. O. Upton, *Glass Ind.* xxi. p. 109 (1940).

‡ A. J. Holland and W. E. S. Turner, *J. Soc. Glass Tech., Trans.* xxv. p. 164 (1941).

I.LXXXI. *The Significance of the Bakerian Lecture of 1843.*

By H. J. J. WINTER, M.Sc. (Lond.), A.Inst.P., University College of the
South-West of England, Exeter *.

[An account of several new Instruments and Processes for determining
the Constants of a Voltaic Circuit. By Sir Charles Wheatstone,
Professor of Experimental Philosophy in King's College, London.
(Phil. Trans. Roy. Soc. cxxxiii. pp. 303-327, 1843) ⁽¹⁾.]

[Received April 5, 1943.]

CONTENTS.

	Page
Introduction	700
Aim of the Researches	701
Acknowledgment to G. S. Ohm	701
Data	701
Use of the Rheostat	702
A Practical Standard of Resistance	704
Determination of Resistance	704
Evaluation of Electromotive Force	705
Resistance of Liquids	706
Wheatstone's Bridge	707
Galvanometers	708
Concluding Remarks	798
References	709

INTRODUCTION.

AMID the stress of the present world upheaval another centenary has passed almost unnoticed, namely, that of the award of the Copley Medal by the Royal Society in 1841 to Dr. G. S. Ohm ⁽²⁾ in recognition of his arduous labours directed toward the elucidation of the mechanism of the electric circuit. Though Ohm's experimental discovery of the well-known law (1826) ⁽³⁾ and his subsequent theoretical treatment (1827) ⁽⁴⁾ have been remembered ⁽⁵⁾, the vast field of modern electrical technology has left little place for that important historical document which first founded a practical technique worthy of the great and simple truth which Ohm had revealed; Sir Charles Wheatstone's lecture of 1843, about which I now write, makes in its generous appreciation of Ohm a suitable means of remembrance of the Copley award to this patient Bavarian investigator ⁽⁶⁾, and through its completeness, directness and simplicity affords perhaps a worthier focus for a centenary paper. It is proposed to review some of its more important physical aspects in their historical relationship.

* Communicated by Professor F. H. Newman,

THE AIM OF THE RESEARCHES.

Wheatstone wrote (§ 1) ⁽⁷⁾ :—

“ I intend in the present communication to give an account of various instruments and processes which I have devised and employed during several years past for the purpose of investigating the laws of electric currents ⁽⁸⁾. The practical object to which my attention has been principally directed, and for which these instruments were originally constructed, was to ascertain the most advantageous conditions for the production of electric effects through circuits of great extent, in order to determine the practicability of communicating signals by means of electric currents to more considerable distances than had hitherto been attempted ” ⁽⁹⁾.

ACKNOWLEDGMENT TO G. S. OHM.

“ In this endeavour, guided by the theory of Ohm and assisted by the instruments I am about to describe, I have completely succeeded. . . . The theory we now possess is amply sufficient to direct us rightly in this inquiry, but experiments have not yet been sufficiently multiplied to enable us to obtain, except in a few cases ⁽¹⁰⁾ the numerical values of the constants which enter into various voltaic circuits; and without this knowledge we can arrive at no accurate conclusions.”

Sir Charles Wheatstone (1802–1875) greatly admired the work of G. S. Ohm (1789–1854), and perceiving that there was an urgent need for accurate practical measurement of resistance (an aspect which Ohm’s researches had left largely unexplored) he applied his extraordinary experimental skill to a method which would be independent of fluctuation of e.m.f. of batteries ⁽¹¹⁾. This improvement of methods of resistance measurement resulted largely from the work of those investigators dealing with the telegraph, of whom Wheatstone was one. He wrote further (§ 2) :—

“ The instruments and processes I am about to describe being all founded on the principles established by Ohm in his theory of the voltaic circuit, and this beautiful and comprehensive theory being not yet generally understood and admitted, even by many persons engaged in original research ⁽¹²⁾, I could scarcely hope to make my descriptions and explanations understood without prefacing them with a short account of the principal results which have been deduced from it.”

THE DATA.

(1) Of these results were the removal of the vague terms intensity and quantity ⁽¹³⁾ and the reduction of the findings of many investigators in all parts of Europe to a few simple and general formulæ.

(2) “ The electromotive force E is the agent which in a closed circuit originates electric current and in an open circuit produces ‘ electroscopic

tension' ⁽¹⁴⁾; the resistance R is numerically equal to the reciprocal of the 'conducting power' which term it supersedes in clarity" ⁽¹⁵⁾.

(3) "When the activity of any portion of the circuit is increased or diminished, either by a change in the electromotive force or in the resistance of that portion, the activity of all the other parts of the circuit increases or decreases in a corresponding degree, so that the same quantity of electricity always passes in the same instant of time through every transverse section of the circuit."

(4) Wheatstone still found it useful to refer to the concept of reduced length—that length of copper wire of a given uniform diameter, the resistance of which was equivalent to the sum of the resistances in the circuit—Ohm being the originator of this concept. For a circuit of two parts of reduced lengths λ , λ^1 connected in parallel, their combined resistance may be represented as $\frac{\lambda\lambda^1}{\lambda+\lambda^1}$.

(5) Sources of electric action differ only in the magnitude of their electromotive forces and their effect is modified solely by the resistance of the circuit.

(6) The electromotive force of any voltaic element depends not upon its dimensions but solely upon the nature of the metals and liquids in contact.

(7) On the above basis, Wheatstone stated the general law

$$F = \frac{nE}{\frac{nRD}{S} + \frac{rl}{s}}$$

where F = "the force" of the current,

n = the number of elements in series, each of electromotive force E ,

R = the specific resistance of the liquid in each element,

D = the thickness of the liquid stratum between the plates,

S = the cross-sectional area of each plate in contact with the liquid,

r = the specific resistance of the connecting wire, of length l and cross-sectional area s .

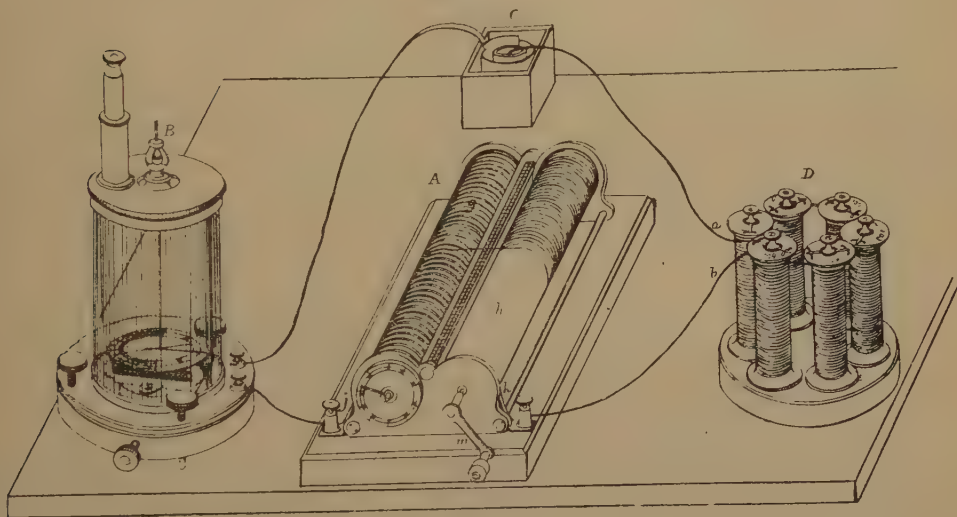
The above was the fullest and most precise statement made on the electric circuit previous to 1845 ⁽¹⁶⁾.

THE USE OF THE RHEOSTAT.

Wheatstone's rheostat marked a great step forward in experimental technique by enabling a current to be maintained constant when fluctuations in electromotive force occurred; this latter factor had been mainly responsible for the tardy acceptance of Ohm's Law in scientific

circles⁽¹⁷⁾. Researches previously carried out with electric circuits depended for their accuracy upon the measurement of current through a galvanometer, *e. g.* “Fechner⁽¹⁸⁾ measured the force of the current by the number of oscillations of the needle when placed at right angles to the coils, a very tedious operation; and others have employed the deviations of the needle, the corresponding degrees of force having been previously determined by some peculiar process, or inferred from some rule depending on the particular construction of the instrument. Another impediment to the use of a galvanometer to measure the force of a current arises from the changes in the magnetic intensity of the needle, which frequently occur, especially when it has been acted upon by too strong a current.” (§ 4.)

Fig. 1.



The principle of the method used earlier by Fechner⁽¹⁹⁾, by Lenz⁽²⁰⁾, and by Pouillet⁽²¹⁾ was demonstrated by Wheatstone thus :—

For a simple circuit $F = \frac{E}{R}$.

Interpose a further resistance R^1 , when “the force of the current” becomes

$$F^1 = \frac{E}{R + R^1}.$$

Therefore,

$$\frac{F}{F^1} = \frac{R + R^1}{R}.$$

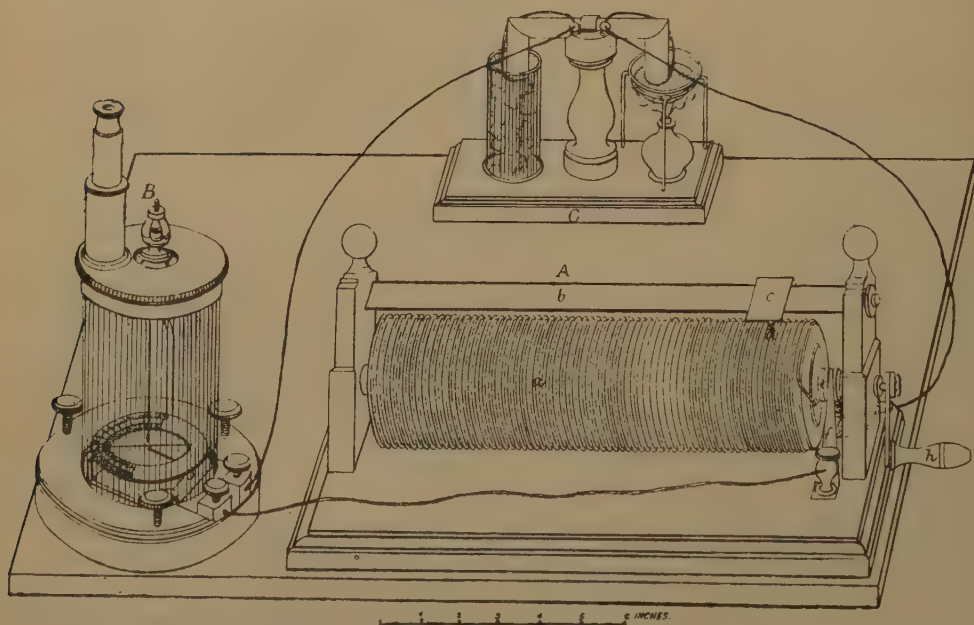
By using a rheostat capable of different low resistance values, Wheatstone removed the difficulty of knowing exactly the forces corresponding to different deviations of the needle. A similar instrument for low-

resistance circuits—the agometer—had been conceived by Jacobi in 1840⁽²²⁾. For high-resistance circuits Wheatstone used resistance coils⁽²³⁾. Two circuit arrangements by Wheatstone are shown in figs. 1 and 2.

A PRACTICAL STANDARD OF RESISTANCE.

Pursuing Ohm's original concept of reduced length, Wheatstone used a practical standard of 1 ft. of copper wire weighing 100 grains. Length and weight were chosen because small changes in diameter produce considerable changes of resistance⁽²⁴⁾. The absolute system of Gauss was in existence⁽²⁵⁾, but its extension to include an absolute unit of resistance came with Wilhelm Weber⁽²⁷⁾. Wheatstone's interests were

Fig. 2.



Two circuits used by Wheatstone, one "hydro-electric," the other thermo-electric.

A. Rheostats.

D. Resistance coils.

essentially practical and there is no indication that he ever thought of anything of an absolute nature. In common with the majority of his contemporaries he failed even to give a name to the units in which he expressed electromotive force, current and resistance⁽²⁷⁾.

DETERMINATION OF RESISTANCE IN A CIRCUIT.

1. Resistance Determined by the Method of Substitution.

In a given electric circuit, replace the unknown resistance R by a suitably adjusted rheostat in order to produce the same deflexion θ in

the galvanometer. Then R equals the resistance of that portion of the rheostat interposed.

2. Resistance of a Galvanometer Coil.

Place a cell of electromotive force E and internal resistance r , a galvanometer of resistance G , and a known resistance R in series.

$$\text{The current} = \frac{E}{R+G+r}.$$

The same current (represented by the same galvanometer deflexion) is now obtained by adding a further identical cell and increasing the resistance by the necessary amount λ by interposing the rheostat.

$$\text{The current} = \frac{2E}{R+G+2r+\lambda}.$$

These expressions lead to

$$G = \lambda - R.$$

3. Internal Resistance of a Cell in terms of Reduced Length.

Six methods, based upon Ohm's Law, are given; Wheatstone's fifth method is outlined below:—

Two identical cells are in series with an external resistance. Using the notation already adopted.

$$\text{The current} = \frac{2E}{R+2r}.$$

Now put the cells "side by side" (*i. e.* in parallel) and introduce resistance λ by means of the rheostat in order to give the same galvanometer deflexion.

$$\text{The current} = \frac{E}{R+r/2+\lambda}.$$

$$\text{Hence} \quad r = R + 2\lambda.$$

In these researches Wheatstone admirably demonstrates the value of the null method. Classical researches in physics invariably show that a theory is commendable in virtue of its simplifying assumptions, and Wheatstone, with his "exactly equal rheomotors" ⁽²⁸⁾, has this same idealism of approach. A recent writer has criticised this outlook ⁽²⁹⁾.

EVALUATION OF ELECTROMOTIVE FORCE.

"The rheostat affords a most ready means of ascertaining the sum of the electromotive forces active in a voltaic circuit, without requiring for this purpose the aid of a rheometer (galvanometer) graduated to indicate proportional forces, or having recourse to the tedious process of counting the oscillations of a needle, employed by Fechner in his investigations." (§ 10.)

With a single cell of electromotive force E and a *total* resistance of R in the circuit, find the extra resistance λ to be introduced to reduce the galvanometer deflexion from 45° to 40° .

Next, with other cells of total electromotive force nE and the same external circuit, find the resistance λ' to be introduced by means of the rheostat (and resistance coils if required) in order to produce the same change in deflexion, 45° to 40° .

Then, in principle,

$$\frac{E}{R+\lambda} = \frac{nE}{nR+n\lambda} = \frac{nE}{nR+\lambda'}$$

Whence $\lambda' = n\lambda$. Knowing the ratio $n\lambda : \lambda$, the ratio $nE : E$ also follows.

Using this method Wheatstone demonstrated experimentally that

(1) the electromotive force of a cell is independent of its dimensions and is determined solely by the nature of the chemicals.

(2) n elements of electromotive force E are equivalent when in series to a total electromotive force of nE .

(3) the "contrary electromotive force" introduced into a circuit by a voltmeter can be measured, using a rheostat: "The measure of this contrary electromotive force is obtained by subtracting the actual number of turns from the number of turns corresponding with the electromotive force of the circuit when the decomposing cell is removed from it." (§ 11.)

Wheatstone's work on electromotive force was extended by Lenz and Saweljew⁽³⁰⁾.

THE RESISTANCE OF LIQUIDS.

Wheatstone laid important foundations in this study. One of his methods involved the use of a column of the liquid in a glass tube "about two inches long and half an inch in internal diameter" and having one fixed and one movable platinum plate:—"I interpose in the circuit a small constant battery, consisting of about three elements, with the rheostat, the resistance coils, the galvanometer, and the measuring tube just described. The end of the piston being a quarter of an inch distant from the fixed plate, I fill the intervening space with the liquid, the resistance of which is to be measured. I then adjust the rheostat to bring the needle of the galvanometer to a determined point; this having been noted, I draw the piston back through the entire remaining space of one inch, and fill the vacancy with the same liquid; the needle will recede towards zero. I then diminish the resistance of the circuit by means of the rheostat and the resistance coils, until the needle stands at the same point that it did when only a quarter of an inch of the liquid column was interposed. The reduced length of the wire thus taken out of the circuit will be the measure of the resistance of one inch of the liquid. The contrary electromotive force arising from the decomposition of the liquid exists in the circuit during the whole process, and therefore does not affect the result." (§ 14.)

For a circuit represented mathematically by $F - E/R$ before interposition of liquid, the resulting modified current F^1 is given by $\frac{E-e}{R+x}$,

where e = the contrary electromotive force of the liquid and x its resistance.

THE WHEATSTONE BRIDGE.

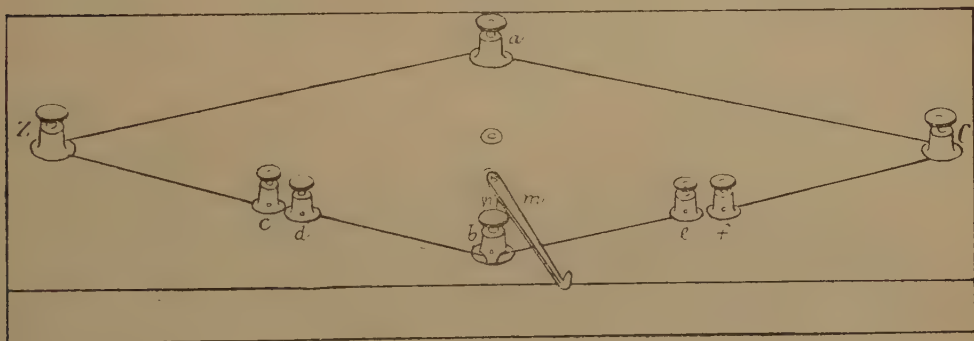
We now come to that part of Wheatstone's work which perpetuates his name in every textbook of electricity, namely, the invention of "the differential resistance measurer" ⁽³¹⁾.

Wheatstone noted that the rheostat was inapplicable when small differences of resistance were to be measured, and further, pointed out that Becquerel's differential galvanometer ⁽³²⁾ was difficult to construct in practice such that currents of equal energy in the two coils would

Fig. 3.



Fig. 4.



Two types of resistance bridge used by Wheatstone.

produce exactly zero deflexion in the needle. An attempt was therefore made to construct a differential instrument having all the advantages of Becquerel's galvanometer but none of its defects. Wheatstone's arrangement also proved to be applicable to any galvanometer.

In the first form of bridge (see fig. 3) four identical copper wires Zb , Za , Ca , Cb , are arranged upon a board such that a cell may be connected from Z to C , and the galvanometer from a to b . The equilibrium of the system is independent of the current (or any fluctuation in it) produced by the cell because equal currents always proceed through the galvanometer in opposite senses. This is due to the precisely equal circuits $ZbaCZ$ and $ZabCZ$.

For the purpose of interposing a known measuring resistance and a resistance whose value is required the wires Zb , Cb are broken from c to d , and from e to f . Thus the disturbance of equilibrium produced by introduction of the unknown resistance can be removed by the introduction of an equal measuring resistance (part of the rheostat known in terms of reduced length).

The second form (fig. 4), has a movable metal arm m able to rest on the wire at a desired point. This form was made more convenient by Kirchhoff⁽³³⁾, who introduced a uniform platinum wire, a movable contact, and a triangular lay-out; the subsequent long, rectangular form was due to Siemens.

OBSERVATIONS RELATING TO GALVANOMETERS⁽³⁴⁾.

1. Wheatstone investigated the influence of the resistance of a galvanometer coil upon the circuit into which it is introduced and developed the theory of the shunt (§ 15).

2. By the theory of branched circuits established by Ohm, Wheatstone found the practical means of ascertaining what division of a galvanometer scale indicates half the current value corresponding to another given reading. In this way, a galvanometer scale may be calibrated, and desired fractions of the main current may be passed through the instrument (§ 18).

3. In § 19 a process of calibration, which superseded those of Nobili⁽³⁵⁾, Melloni⁽³⁶⁾, and Becquerel⁽³⁷⁾, was set down:—

- (i) Determine the total resistance of the circuit when the needle reads 1° .
- (ii) By means of the rheostat and resistance coils reduce the resistance to one half. Then the "force of the current" will be doubled.
- (iii) Continue to reduce the resistance to one-third, one-quarter, etc., and the corresponding "force of the current" will be three, four, etc.
- (iv) In general, if reduced lengths a , b , c , etc., must be removed from the circuit in order to advance the deflexion by 1° at each step, then the forces corresponding to each successive degree are

$$\frac{1}{R}, \frac{1}{R-a}, \frac{1}{R-(a+b)}, \frac{1}{R-(a+b+c)}, \text{ etc.}$$

CONCLUDING REMARKS.

(1) The electrical researches of G. S. Ohm had not received unanimous support because succeeding investigators had not all been able to verify Ohm's Law with certainty. They had used cells (with consequent difficulties due to polarization) whereas Ohm had used thermoelectric elements. Nor were methods for current measurement and current regulation sufficiently reliable. The rheostat of Wheatstone and his

knowledge of galvanometry enabled accurate practical researches based on Ohm's Law to be made, and laid a basis for D.C. technology.

(2) Wheatstone made the first really clear statement in English of Ohm's Law and its associated definitions⁽³⁸⁾. The effect of resistance and the factors determining electromotive force are explicitly stated, and the behaviour of a circuit containing liquid is expressed mathematically for the first time.

(3) The advances made by Wheatstone were essentially those of practical technique. His work suggested no generality respecting branched circuits, nor any absolute standard of reference in resistance measurement. The value of null methods in galvanometry was stressed.

(4) Wheatstone's Bakerian Lecture was reproduced in Poggendorff's *Annalen* (vol. lxii., pp. 499-543 (1844)), and his work had its influence upon contemporary physicists on the Continent, in particular on Professor M. Jacobi at St. Petersburg, and his countrymen Lenz and Saweljew. The place of Wheatstone's work in the complete historical development of Ohm's Law has been treated elsewhere⁽³⁹⁾.

In conclusion, I thank Professor F. H. Newman, Vice-Principal of University College, Exeter, whose help and encouragement I much appreciate; also Miss K. E. Perrin, Acting Librarian of the same College, who has kindly permitted to me very free use of the Library's facilities, and Mr. A. J. Ansley, to whom I am further indebted.

REFERENCES.

- (1) See also "The Scientific Papers of Sir Charles Wheatstone," Phys. Soc. Lond. pp. 97-132 (1879).
- (2) Proc. Roy. Soc. iv. p. 336 (1841).
- (3) G. S. Ohm, Schweigger's 'Journal,' Bd. 46, pp. 137-166 (1826).
- (4) G. S. Ohm, 'Die galvanische Kette mathematisch bearbeitet.' T. H. Riemann: Berlin, 1827. Also in Ohm's 'Gesammelte Abhandlungen' (introduction by E. Lommel), pp. 61-187. Leipzig, 1889.
- (5) Aus Georg Simon Ohm's 'Handschriftlichem Nachlass.' L. Hartmann: Munich, 1927. See also 'Nature,' Feb. 14th, 1889.
- (6) G. S. Ohm was born on 16th March, 1789, at Erlangen, in Bavaria. His researches upon the electric circuit were performed under considerable difficulty during his spare time whilst a teacher at the Jesuit Gymnasium in Cologne, 1817-1826. Leave of absence during the year 1827 enabled the theoretical treatment (see (4) above) to be completed in Berlin.
- (7) § refers to the sections of the Wheatstone lecture under discussion.
- (8) The first verification and acknowledgment of Ohm's Law was by G. T. Fechner: 'Mass bestimmungen über die galvanische Kette.' F. A. Brockhaus: Leipzig, 1831. In 1837 independent work by the French investigator Pouillet, who first successfully used the tangent and sine galvanometers, also established the validity of Ohm's Law, and Ohm, in a letter written to Ludwig I. of Bavaria on 6th May, 1842, said that he did not expect Professor Pouillet "to adorn himself with foreign feathers." In fact, Pouillet was given considerable praise in some French treatises at the expense of Ohm, e. g., M. J. Jamin wrote about "le remarquable travail d'un savant français, Pouillet, qui, à l'aide du galvanomètre, a retrouvé la loi de Ohm, encore inconnue en France

en 1837, et l'a fait définitivement adopter par les physiciens de tous les pays," *vide* Pouillet, 'Comptes Rendus,' iv. pp. 267-279 (1837)/ Letter to Ludwig I. "Akten des Bayerischen Staatsministerium, für Unterricht und Kultus München."/M. J. Jamin: 'Cours de Physique de l'Ecole Polytechnique, iv. pt. 1, p. 30 (1888).

- (9) As a result there arose the misconception about the velocity of electricity.
- (10) Reliable experimental work involving Ohm's Law had been made by P. S. Munck af Rosenschöld in Lund (Pogg. *Ann.* Bd. 43, pp. 193-228, pp. 440-493 (1838)); Vorrsselmann de Heer in Holland (Pogg. *Ann.* Bd. 46, pp. 513-537 (1839)); E. Lenz (Pogg. *Ann.* Bd. 34, pp. 418-437 (1835)); and M. Jacobi (Pogg. *Ann.* Bd. 48, pp. 26-58 (1839)) in Russia; and by Pouillet (Pogg. *Ann.* Bd. 42, pp. 281-297 (1837)) in France as mentioned. Researches by the last three investigators are most fruitful of experimental values. See also the letter to Ludwig I.
- (11) Ohm actually determined his Law using thermoelectricity, and therefore with greater precision than batteries would at that time allow. Poggendorff suggested to Ohm the use of thermojunctions on the basis of Seebeck's discovery (1821). Antoine César Becquerel had devised a differential galvanometer in which any fluctuation of current produced equal and opposite magnetic fields at the needle. (*Annales de Chimie*, xxxii. p. 423 (1826).
- (12) The main difficulty was polarization. Ohm's views were upheld by his countrymen Schweigger, Poggendorff, Pfaff, and Fechner, by Munck af Rosenschöld and Vorrsselmann de Heer, and by Joseph Henry. Amongst his critics was the Hegelist G. F. Pohl in 'Jahrbücher für wissenschaftliche Kritik,' Nos. 11-14 (1828). Ohm has written: "Pohl is well-known to be arrogant and his blindness in despising my work is only due to his own attempt to restrain me. He is misguided his own animosity and is not led by truth." (Letter by G. S. Ohm to Prussian Ministry in Berlin, 6th March, 1828. No. 669 in 'Urkundensammlung des Deutschen Museums München.')
- (13) See Ronalds, *Phil. Mag.* xlv. p. 261 (1815).
- (14) The electroscopic tension mentioned by Ohm remained ill-defined until G. Kirchhoff (Pogg. *Ann.* Bd. 78 (1849)) identified it with electrostatic potential.
- (15) The term "conducting power" had been used by the earlier investigators of electrical conductivity. (A. C. Becquerel, *Annales de Chimie*, xxxii. p. 423 (1826); H. Davy, *Phil. Trans.* cxi. p. 425 (1821)).
- (16) A fair account existed in German: see P. O. C. Vorrsselman de Heer (ref. 10 above).
- (17) See letter from J. C. Poggendorff to G. S. Ohm, dated 26th April, 1844. No. 735 in 'Urkundensammlung des Deutschen Museums München.' The controversy relating to the origin of the electric current in the voltaic cell (contact hypothesis versus chemical hypothesis), together with the widespread faith in Hegel's philosophy in the German universities, were also contributory.
- (18) See G. T. Fechner, 'Massbestimmungen über die galvanische Kette,' p. 5. Leipzig (1831). The principal types of galvanometer used previous to 1843 depended upon—
 - (i) Return of a deflected magnetic needle by a torsion-head (Ohm). (A bifilar suspension was not used.)
 - (ii) Oscillation of a magnetic needle having the rest position of its axis at right-angles to the magnetic field due to the current (Fechner).
 - (iii) Equilibrium of a magnetic needle hanging freely under gravity when submitted to action of the field due to current (Becquerel, improved by Lenz and Jacobi).

- (iv) Measurement of the impulse due to momentary current (Lenz).
- (v) Steady deflexion measured for a magnetic needle (uni- or bifilar suspension) as in a tangent galvanometer or similar instrument (Gauss, Nervander, Pouillet, Jacobi, Lenz.)
- (vi) Alteration of the field direction in order to keep the needle at a definite position (sine galvanometer of Pouillet).
- (19) 'Massbestimmungen über die galvanische Kette,' pp. 28-31.
- (20) Lenz, *Pogg. Ann.* Bd. 48, p. 393 (1839).
- (21) Paper referred to in (8) above.
- (22) Footnote to § 4. Also see 'Athenaum,' No. 678 (1840). Wheatstone and Jacobi met in England in Aug. 1840 and exchanged ideas. Jacobi repeated Wheatstone's method on his return to Russia (*Pogg. Ann.* Bd. 54, 1841; Bd. 62 (1842)).
- (23) The resistance box was first used by Werner Siemens (see 'Die Lehre vom Galvanismus.' G. Wiedemann. 2nd edition, Bk. 1, p. 239. Braunschweig (1872)).
- (24) Previous to the conception of an absolute resistance standard each investigator used his own arbitrary standard, *e. g.* Lenz, "The resistance of 1 ft. no. 11 copper wire" (*Pogg. Ann.* Bd. 34, p. 418 (1833)). Vosselman de Heer, "The resistance of 1 m. of Copper wire of 1 mm. diameter" (*Pogg. Ann.* Bd. 46, pp. 513-537 (1839)).
- (25) K. F. Gauss, 'Intensitas vis magneticae terrestres ad mensuram absolutam revocata.' Göttingen, 1833.
- (26) W. Weber, *Pogg. Ann.* Bd. 82, p. 337 (1851).
- (27) "When the Committee was first appointed no coherent system of units for the measurement of electrical resistance, current, electromotive force, quantity, or capacity, had met with general approval. It was true that Professor W. Weber's absolute system existed on paper, but it was not understood or used by practical men." (Introduction to Reports of the Committee on Electrical Standards appointed by the British Association; Cambridge, 1913. This Committee first met in 1861 under W. Thomson (later Lord Kelvin).)
- (28) The word "rheostat" is the only survival of a system of nomenclature (in which a cell was called a rheomotor) proposed by Wheatstone in this lecture.
- (29) N. Campbell, *Proc. Phys. Soc.* xlviii. p. 708 (1936).
- (30) Lenz and Saweljew, *Pogg. Ann.* Bd. 67, pp. 497-528 (1846).
- (31) The principle had been proposed by Hunter Christie (*Phil. Trans.* 1833), to whom Wheatstone made acknowledgment.
- (32) See ref. (11) above.
- (33) G. Kirchhoff, 'Gesammelte Abhandlungen,' p. 15. Leipzig (1882). (From *Pogg. Ann.* Bd. 64 (1845).)
- (34) Difficulties due to galvanometers had been stressed by G. S. Ohm (*Schweigger's Journ.* Bd. 63, pp. 1-26, 159-189 (1831), and the varied views of Jäger, Becquerel, Fechner, De la Rive, Nobili, Davy, Walker, Berzelius, Parrot, and Ritter were discussed.
- (35) L. Nobili, 'Antologia di Firenze,' No. 142 (1832).
- (36) Nobili and Melloni, *Pogg. Ann.* Bd. 27, pp. 439-455 (1833).
- (37) A. C. Becquerel. See reference (11). (References 36, 37, 38 are not exhaustive.)
- (38) 'Die galvanische Kette mathematisch bearbeitet' had appeared in English in 1841 in Richard Taylor's *Scientific Memoirs*, Vol. II. (Translation by Dr. W. Francis.) Ohm's exposition, like that of Fechner's, was often, however, tedious.
- (39) H. J. J. Winter, M.Sc., Dissertation, University of London, 1940, from which this paper has been developed,

LXXXII. *Note on the Interaction of Two Point-Charges.*

By E. A. MILNE, F.R.S.*

[Received July 26, 1943.]

1. THE object of the present note is to give a simplified version of the derivation of the integrals of "energy" and "angular momentum" for a pair of point-charges in one another's presence, as found in a recent series of papers†. These integrals were of considerable interest in that they contained explicitly factors involving the "classical radius of the electron," and so allowed the calculation of the nature of the interaction when the two particles were at a distance comparable with this length.

In the series of papers mentioned, the deductive theory of electromagnetism developed in 1938‡ was applied to the motion of a particle of charge e_1 and mass m in the presence of a relatively massive particle of charge e_2 . The massive particle could be considered as permanently coincident with a fundamental particle of the substratum. This could be taken as origin, and as the seat of the observer making the measurements in terms of which the calculations were conducted. The investigation was thus carried out in terms of t -measures, *i. e.*, measures based on the kinematic or t -scale of time. This is the most appropriate scale to use in a fundamental investigation, as it can be employed most conveniently with ordinary flat Euclidean space, the private space of the observer. The equations were then transformed into the more familiar dynamical or τ -scale, which is locally, and for epochs close to the present, equivalent to our ordinary code of measurement as used in empirical physics.

But it may be a convenience to some readers if the calculations embarked on in Part IV. of the series of papers are re-formulated directly in terms of τ -measures and the τ -scale of time. It will thus be easier to link the new ideas with the classical ideas, which of course were necessarily formulated in the τ -scale. It will also become easier to see how the specific features in the new integrals of "energy" and "angular momentum" take their origin.

2. In what follows, I make no apology for going into elementary mathematical details. The complaint is so often made nowadays by physicists that they cannot follow the work of theoretical investigators.

* Communicated by the Author.

† "Rational Electrodynamics, I-V.," *Phil. Mag.* xxxiv. p. 73 (Feb. 1943); xxxiv. p. 197 (March 1943); xxxiv. p. 325 (April 1943).

‡ *Proc. Roy. Soc.* 165 A, pp. 313 and 333 (1938).

that when an opportunity occurs of exhibiting calculations in a form which could readily be reproduced in a text-book, it seems as well to take it.

3. *Equation of motion.*—As stated, we consider a massive particle, of charge e_2 , at a fundamental particle which we take to be the origin O. The observer at O assigns at epoch τ , measured on the dynamical time-scale, a vector distance \mathbf{r} to a second particle, of charge e_1 and mass m . The rational electrodynamics which was developed by starting with charged particles in motion showed that when all relativistic considerations are taken into account, the electric intensity \mathbf{E}_1 , exerted at e_1 due to e_2 , is given in τ -measures by the relation

$$\mathbf{E}_1 = \frac{e_2}{(1-v^2/c^2)^{\frac{1}{2}}} \frac{\mathbf{r}}{|\mathbf{r}|^3}, \quad \dots \dots \dots (1)$$

where v , the speed of the particle at \mathbf{r} , is given by

$$v^2 = \left(\frac{d\mathbf{r}}{d\tau}\right)^2 = \left(\frac{dr}{d\tau}\right)^2 + r^2 \left(\frac{d\theta}{d\tau}\right)^2, \quad \dots \dots \dots (2)$$

r denoting $|\mathbf{r}|$. Formula (1) exhibits \mathbf{E}_1 as an inverse-square, Coulomb force, modified by a relativistic factor $(1-v^2/c^2)^{-\frac{1}{2}}$. It is unwise to attempt to seek a simple argument for the presence of this dynamical factor—it originated in fundamental considerations arising from the “equivalence” of the particle observers who constitute the members of the substratum in the t -dynamics.

The magnetic intensity \mathbf{H}_1 at e_1 is given similarly by the relation

$$\mathbf{H}_1 = \frac{e_2}{(1-v^2/c^2)^{\frac{1}{2}}} \frac{\frac{1}{2}(0+\dot{\mathbf{r}})}{c} \wedge \frac{\mathbf{r}}{|\mathbf{r}|^3}. \quad \dots \dots \dots (3)$$

We write \mathbf{H}_1 in the form (3) to show explicitly that we have taken proper account of the circumstance that the magnetic field originates from the interaction of a particle at rest relative to the observer, with a particle of velocity $\dot{\mathbf{r}}$ relative to the observer. It will be remembered that one of the features of the new theory of electromagnetism, arising from a proper incorporation of the relativity of motion into the treatment of magnetic intensity, was the replacement of the usual Biot and Savart law by a law involving the arithmetic vector-mean of the velocities of the test-particle at r and the field-originating particle, both relative to the observer. In the present case * this is $\frac{1}{2}(0+\dot{\mathbf{r}})$, *i. e.* just $\frac{1}{2}\dot{\mathbf{r}}$.

The Larmor-Lorentz pondero-motive force due to the pair $(\mathbf{H}_1, \mathbf{E}_1)$ acting on e_1 is given by the expression

$$e_1 \left[\mathbf{E}_1 + \frac{\mathbf{v} \wedge \mathbf{H}_1}{c} \right], \quad \dots \dots \dots (4)$$

$$\text{i. e.,} \quad \frac{e_1 e_2}{(1-v^2/c^2)^{\frac{1}{2}}} \left[\frac{\mathbf{r} + \frac{1}{2}\dot{\mathbf{r}} \wedge (\dot{\mathbf{r}} \wedge \mathbf{r})/c^2}{|\mathbf{r}|^3} \right]. \quad \dots \dots \dots (5)$$

* This particular case of the general theory was first found by L. H. Thomas, *Phil. Mag.* (7) iii. p. 1 (1927).

It follows that when we use *Einstein's mechanics*, the equation of motion of the particle of mass m at \mathbf{r} is

$$\frac{d}{d\tau} \left[\frac{m\dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} \right] = \frac{e_1 e_2}{(1-v^2/c^2)^{\frac{1}{2}}} \frac{\mathbf{r} + \frac{1}{2}(\mathbf{r} \wedge \dot{\mathbf{r}}) \wedge \dot{\mathbf{r}}/c^2}{|\mathbf{r}|^3}. \quad (6)$$

In "Rational Electrodynamics" I did not use Einstein's mechanics, and the work involved more complicated sets of factors $(1-v^2/c^2)^{-\frac{1}{2}}$ in its strictly relativistic dress. I believe the t -treatment in "R.E." to be "exact," but when passing from t -measures to τ -measures I made certain approximations. It seems simpler in the present paper to start with Einstein's mechanics and the Larmor-Lorentz form of the mechanical force, but I do not claim that (6) is rigorously "exact" in the sense that the original t -equations were "exact." Equation (6), however, has the same integrals as the original t -equations, when transformed to τ -measure, to a very close approximation. The present treatment has the advantage of having a more familiar structure than the corresponding t -equations; and it will show how the two integrals, in τ -measure, arise respectively from the electric and magnetic terms on the right-hand side of (6).

4. "*Energy*"-integral.—Multiply (6) scalarly by the vector $\dot{\mathbf{r}}$. Since a triple product of vectors vanishes when two factors are alike, we get

$$(1-v^2/c^2)^{\frac{1}{2}} \dot{\mathbf{r}} \cdot \frac{d}{d\tau} \left[\frac{m\dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} \right] = e_1 e_2 \frac{\mathbf{r} \cdot \dot{\mathbf{r}}}{|\mathbf{r}|^3}. \quad (7)$$

The right-hand side is just

$$-e_1 e_2 \frac{d}{d\tau} \left(\frac{1}{|\mathbf{r}|} \right).$$

The left-hand side may be written in the form

$$mc^2(1-v^2/c^2) \left[\frac{\mathbf{r}/c}{(1-v^2/c^2)^{\frac{1}{2}}} \cdot \frac{d}{d\tau} \frac{\dot{\mathbf{r}}/c}{(1-v^2/c^2)^{\frac{1}{2}}} \right],$$

or

$$\frac{1}{2} mc^2 (1-v^2/c^2) \frac{d}{d\tau} \left[\frac{\dot{\mathbf{r}}/c}{(1-v^2/c^2)^{\frac{1}{2}}} \right]^2.$$

Since $\dot{\mathbf{r}}^2 = v^2$, this may be re-written as

$$\frac{1}{2} mc^2 (1-v^2/c^2) \frac{d}{d\tau} \left[-1 + \frac{1}{1-v^2/c^2} \right],$$

which is just

$$mc^2 \frac{d}{d\tau} \log \frac{1}{(1-v^2/c^2)^{\frac{1}{2}}}.$$

Hence (7) becomes

$$\frac{d}{d\tau} \log \frac{1}{(1-v^2/c^2)^{\frac{1}{2}}} = -\frac{e_1 e_2}{mc^2} \frac{d}{d\tau} \left(\frac{1}{|\mathbf{r}|} \right).$$

The integral of this is

$$\log \frac{1}{(1-v^2/c^2)^{\frac{1}{2}}} + \frac{e_1 e_2}{mc^2} \frac{1}{|\mathbf{r}|} = \text{const.} = \log \left(1 + \frac{W}{mc^2} \right),$$

say, where $W > -mc^2$ and W is constant. This gives

$$\frac{1}{(1-v^2/c^2)^{\frac{1}{2}}} = \left(1 + \frac{W}{mc^2}\right) e^{-\frac{e_1 e_2}{mc^2} \frac{1}{r}} \quad (8)$$

This is the desired "energy-integral" (relation (101) of "R.E.").

5. "Angular momentum"-integral.—Multiply (6) vectorially by \mathbf{r} . We get

$$\mathbf{r} \wedge \frac{d}{d\tau} \left[\frac{\dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} \right] = \frac{\frac{1}{2} e_1 e_2}{mc^2} \frac{\mathbf{r} \wedge [(\mathbf{r} \wedge \dot{\mathbf{r}}) \wedge \dot{\mathbf{r}}]}{|\mathbf{r}|^3 (1-v^2/c^2)^{\frac{1}{2}}} \quad (9)$$

The left-hand side is just

$$\frac{d}{d\tau} \left[\frac{\mathbf{r} \wedge \dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} \right].$$

The right-hand side is, on expanding the continued vector product,

$$\begin{aligned} & \frac{\frac{1}{2} e_1 e_2}{mc^2} \frac{\mathbf{r} \wedge [-\mathbf{r}(\dot{\mathbf{r}}^2) + \mathbf{r}(\mathbf{r} \cdot \dot{\mathbf{r}})]}{|\mathbf{r}|^3 (1-v^2/c^2)^{\frac{1}{2}}}, \\ i. e. & -\frac{1}{2} \frac{e_1 e_2}{mc^2} \frac{\mathbf{r} \wedge \dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} \frac{d}{d\tau} \left(\frac{1}{|\mathbf{r}|} \right). \end{aligned}$$

Thus (9) becomes

$$\frac{d}{d\tau} \left[\frac{\mathbf{r} \wedge \dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} \right] + \frac{1}{2} \frac{e_1 e_2}{mc^2} \frac{\mathbf{r} \wedge \dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} \frac{d}{d\tau} \left(\frac{1}{|\mathbf{r}|} \right) = 0.$$

The integral of this is

$$\frac{\mathbf{r} \wedge \dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} e^{\frac{1}{2} \frac{e_1 e_2}{mc^2} \frac{1}{|\mathbf{r}|}} = \mathbf{a} \text{ a constant vector}$$

Hence

$$\frac{\mathbf{r} \wedge \dot{\mathbf{r}}}{(1-v^2/c^2)^{\frac{1}{2}}} = \mathbf{C} e^{-\frac{1}{2} \frac{e_1 e_2}{mc^2} \frac{1}{r}} \quad (10)$$

Multiplying (10) scalarly by \mathbf{r} , we have

$$\mathbf{C} \cdot \mathbf{r} = 0,$$

whence \mathbf{r} is always perpendicular to a constant vector. Hence the orbit lies in a plane. It is therefore sufficient to replace the vector equation (10) by its scalar equivalent. Using polar variables this gives

$$\frac{mr^2 d\theta/d\tau}{(1-v^2/c^2)^{\frac{1}{2}}} = H e^{-\frac{1}{2} \frac{e_1 e_2}{mc^2} \frac{1}{r}}, \quad (11)$$

where H is a constant of the dimensions of angular momentum. Equation (11) is the desired integral of "angular momentum" (relation (112) of "R.E.").

6. To make this paper self-contained, it is worth while deducing the equation of the orbit from the integrals (8) and (11). Dividing (11) by (8), we have

$$mr^2 \frac{d\theta}{d\tau} = \frac{H}{1+W/mc^2} e^{\frac{1}{2}x}, \quad (12)$$

where
$$x = \frac{e_1 e_2}{mc^2} \frac{1}{r} \quad \dots \quad (13)$$

But, by (2) and (8),

$$1 - \frac{v^2}{c^2} = 1 - \frac{1}{c^2} \left(\frac{d\theta}{d\tau} \right)^2 \left[r^2 + \left(\frac{dr}{d\theta} \right)^2 \right] = \left(1 + \frac{W}{mc^2} \right)^{-2} e^{2x}. \quad (14)$$

Substituting for $d\theta/d\tau$ from (12) in (14) so as to eliminate the time, we have

$$1 - \frac{H^2/m^2 c^2}{(1 + W/mc^2)^2} \frac{e^x}{r^4} \left[r^2 + \left(\frac{dr}{d\theta} \right)^2 \right] = \frac{e^{2x}}{(1 + W/mc^2)^2}.$$

Replacing the variable r by the variable x from (13), this becomes

$$\left(1 + \frac{W}{mc^2} \right)^2 e^{-x} - e^x = \frac{H^2 c^2}{e_1^2 e_2^2} \left[x^2 + \left(\frac{dx}{d\theta} \right)^2 \right],$$

or
$$x^2 + \left(\frac{dx}{d\theta} \right)^2 = \alpha^2 \left[\left(1 + \frac{W}{mc^2} \right)^2 e^{-x} - e^x \right], \quad \dots \quad (15)$$

where α (the fine-structure constant for this system) is given by

$$\alpha = \frac{e_1 e_2}{Hc^2}. \quad \dots \quad (16)$$

7. The further discussion of (8), (11) and (15) has been given in Parts IV. and V. of "R E." We may emphasize here that the present simple treatment has sufficed to throw up the length $e_1 e_2 / mc^2$, the effective "radius" of the moving point-charge, and α , the "fine-structure" constant of the system, by classical methods, using classical relativistic dynamics but the new modified electromagnetic theory. The integrals (8) and (11), with orbital equation (15), were, however, first obtained using the deductive or t -dynamics. The present simpler deduction, on a less fundamental basis, is not intended to replace the earlier work, but to exhibit the interesting fact that the existence of integrals (8) and (11) depends intimately on the presence of the factor $(1 - v^2/c^2)^{-\frac{1}{2}}$ in \mathbf{E}_1 and \mathbf{H}_1 and on the way in which the velocity of the moving particle contributes to the form of \mathbf{H}_1 .

LXXXIII. Distribution in Time of Spontaneous Fluctuation Voltage.

By M. SURDIN, D.Sc.*

[Received March 15, 1943.]

Introduction.

SPONTANEOUS fluctuation voltage is understood to be such fluctuation voltages as are due to thermal-agitation in resistors, shot-effect, flicker and resistance fluctuation effects⁽¹⁾. During the last two decades extensive

* Communicated by the Author.

experimental and theoretical work was carried out, and chiefly because of the study of the first two phenomena led to formulæ giving the mean-square of the fluctuation voltage or current. However, for the better understanding of these phenomena and for certain practical applications, a knowledge of the distribution in time of the fluctuation voltage is desirable.

It is proposed to consider this problem from the thermodynamical point of view, since it has the greatest general application and in the case of voltage fluctuations avoids the use of the somewhat difficult concept of "random phase." The derivation given here, of the probability of distribution of a variable describing a closed system, is somewhat different from the one given by Einstein in his general treatment of fluctuations ⁽²⁾, and follows closely Landau and Lifschitz's ⁽³⁾.

Consider a closed system in which x is some physical quantity characteristic of the state of the system. Then the problem is to determine the probability that this quantity has a value x different from the mean value x_0 which it has in the state of equilibrium.

Let $d\Gamma$ be the element of phase space of the system corresponding to values of x between x and $x+dx$. The probability, dp , that the system is found in this volume, which is equivalent to the probability that x has the value in the given range dx , is proportional to the volume $d\Gamma$ of this element, i. e.

$$dp \propto d\Gamma. \quad (1)$$

$$\text{Let} \quad \Delta\Gamma = e^S, \quad (2)$$

where S is the entropy and $\Delta\Gamma$ the statistical weight of the macroscopic state for which the mean value of the quantity concerned is x . Then

$$dp \propto e^S \cdot \frac{d\Gamma}{\Delta\Gamma}. \quad (3)$$

But as the volume of any region of phase space is proportional to the corresponding range of values of x , it is apparent that

$$\frac{d\Gamma}{\Delta\Gamma} = \frac{dx}{\Delta x}. \quad (4)$$

$$\text{Thus} \quad dp \propto e^S \cdot dx/\Delta x. \quad (5)$$

Both e^S and $1/\Delta x$ are functions of x , but the dependence of e^S on x being an exponential is the most important, and so $1/\Delta x$ may be considered as constant, and dp as proportional to e^S alone, so that

$$dp \propto e^S \cdot dx. \quad (6)$$

Expanding S in the region x_0 ,

$$S = S(x_0) - \frac{\lambda}{2}(x-x_0)^2 + \dots, \quad (7)$$

$$\text{where} \quad \lambda = \left(\frac{\partial^2 S}{\partial x^2} \right)_{x=x_0}; \quad (8)$$

λ is a positive quantity because the entropy of a closed system is a maximum for a state of statistical equilibrium, thus

$$\left(\frac{\partial S}{\partial x}\right)_{x=x_0} = 0 \quad \text{and} \quad \left(\frac{\partial^2 S}{\partial x^2}\right)_{x=x_0} < 0, \quad (9)$$

so that
$$dp = A \cdot \exp \left[-\frac{\lambda}{2} (x - x_0)^2 \right] dx, \quad (10)$$

and
$$\overline{(x - x_0)^2} = \frac{\int_{-\infty}^{+\infty} \exp \left[-\frac{\lambda}{2} (x - x_0)^2 \right] \cdot (x - x_0)^2 \cdot d(x - x_0)}{\int_{-\infty}^{+\infty} \exp \left[-\frac{\lambda}{2} (x - x_0)^2 \right] \cdot d(x - x_0)} = \frac{1}{\lambda}. \quad (11)$$

The integration is performed over all possible values of $x - x_0$. The inaccuracy of carrying out the integration, in ranges where $x - x_0$ is large, is not great, since the contribution to the total integrals will in any case be small in these ranges. Hence

$$dp = A \cdot \exp \left[-\frac{1}{2} \frac{(x - x_0)^2}{\overline{(x - x_0)^2}} \right] dx. \quad (12)$$

The constant A is determined by the normalizing condition

$$A \int_{-\infty}^{+\infty} \exp \left[-\frac{1}{2} \frac{(x - x_0)^2}{\overline{(x - x_0)^2}} \right] dx = 1, \quad (13)$$

so that finally

$$dp = [2\pi \overline{(x - x_0)^2}]^{-\frac{1}{2}} \cdot \exp \left[-\frac{1}{2} \frac{(x - x_0)^2}{\overline{(x - x_0)^2}} \right] dx. \quad (14)$$

Application to Voltage Fluctuations.

Consider an amplifier, the band width of which is $\Delta\nu$, and let v be the instantaneous output voltage due to fluctuations, $\overline{E^2}$ be the mean-square value of v . If $\overline{e_v^2}$ is the spectrum density of v^2 at the frequency ν , then by definition

$$\overline{E^2} = \int_{\nu_1}^{\nu_n} \overline{e_v^2} \cdot d\nu, \quad (15)$$

where ν_1 and ν_n are the limiting frequencies of the amplifier. In the case of thermal fluctuations

$$\overline{E^2} = 4RkT\Delta\nu, \quad (16)$$

where R is the equivalent input resistance, T its absolute temperature and k is Boltzmann's constant.

Consider now the amplifier and its supply voltages to be thermally insulated, and let the applied input voltage be zero, then $v_0 = \bar{v} = 0$. Applying Einstein's results to the amplifier, characterized by the voltage v , the probability of finding v at any time in the range dv is

$$dp = (2\pi \overline{E^2})^{-\frac{1}{2}} \cdot \exp \left(-\frac{1}{2} \frac{v^2}{\overline{E^2}} \right) dv. \quad (17)$$

The probability that the fluctuation voltage will lie between zero and v at any given instant is

$$p = (2\pi\overline{E^2})^{-\frac{1}{2}} \int_0^v \exp\left[-\frac{1}{2} \frac{v^2}{\overline{E^2}}\right] dv, \quad (18)$$

and the probability that the voltage at any given instant will exceed v , is

$$q = (2\pi \cdot \overline{E^2})^{-\frac{1}{2}} \int_v^\infty \exp\left[-\frac{1}{2} \frac{v^2}{\overline{E^2}}\right] dv. \quad (19)$$

To derive equation (17), no assumption has been made as to the ratio of the inverse of the "time of correlation" to the highest frequency passed by the amplifier. Eq. (17) is thus valid when e_v^2 is a function of the frequency, in particular in the case of shot-effect, for frequencies of the order, or higher than the inverse of the time of flight of electrons*. Inversely, it may be proved (see Appendix I.), that if the probability of finding v in the range dv follows the "normal law," it is valid for all types of fluctuation whatever the ratio of the inverse of the "time of correlation" to the highest frequency passed by the amplifier.

The mean absolute value of v is

$$|\overline{v}| = 2(2\pi\overline{E^2})^{-\frac{1}{2}} \int_0^\infty v \exp\left[-\frac{1}{2} \frac{v^2}{\overline{E^2}}\right] dv = \left[\frac{2}{\pi} \overline{E^2}\right]^{\frac{1}{2}} \simeq 0.8 \sqrt{\overline{E^2}}. \quad (20)$$

An interesting distribution is that of v^2 . It can be seen that it does not follow the "normal law," in fact, re-writing eq. (17) in the form

$$dp = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} y^2\right) dy, \quad (21)$$

where $y = v/\sqrt{\overline{E^2}}$, and let $z = y^2$, then the probability of finding z in the range dz is

$$d\omega = (2\pi)^{-\frac{1}{2}} z^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} z\right) dz. \quad (22)$$

The mean value of z (z having the range of variation from 0 to ∞), is

$$\bar{z} = (2\pi)^{-\frac{1}{2}} \int_0^\infty z^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} z\right) dz = 1, \quad (23)$$

as would be expected. Eq. (23) justifies experimental methods, using a square-law detector or thermocouple for the measurement of $\overline{E^2}$.

The responses of a linear and a square-law detector may now be compared. The probability of distribution of fluctuation voltages, after linear rectification (no smoothing), will remain a "normal law," whereas after a square-law rectification the distribution will be given by eq. (22). It will be seen that a linear detector is more advantageous; the figure of merit, if taken as the mean fluctuation voltage after rectification, is 0.8 for a full-wave linear rectifier and 1 for a full-wave square-law rectifier.

* Compare with Landon⁽⁴⁾, who proved a similar law for fluctuations satisfying the condition of "random phase," i. e. when the inverse of the "time of correlation" is large compared to the highest frequency passed by the amplifier.

Appendix I.

To prove that when the probability of distribution of the instantaneous voltage fluctuation follows the "normal law" it is valid in cases of fluctuations, where $\overline{e_v^2}$ is a function of the frequency ν , the additive property of the "normal law" is used. That is to say, if $x = \sum_i x_i$ and $\overline{x_i^2} = a_i^2$; $\overline{x^2} = \sum_i \overline{x_i^2} = a^2$, and if the probability of finding x_i in the range dx_i is

$$dp_i = (2\pi a_i^2)^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} \frac{x_i^2}{a_i^2}\right) dx_i, \quad \dots \quad (A1)$$

the probability dp of finding x in the range dx is

$$dp = (2\pi a^2)^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} \frac{x^2}{a^2}\right) dx. \quad \dots \quad (A2)$$

Consider now a large number of amplifiers, each having a flat-topped response curve centred on the frequencies $\nu_1, \nu_2, \dots, \nu_n$ and having a band-width of $\Delta\nu_i$, all amplifiers having the same gain and with their inputs and outputs connected in series. Let $\overline{e_{\nu_i}^2}$ be the spectrum density of the noise at the frequency ν_i , then, considering the i^{th} amplifier, the probability of finding v_i in the range dv_i is

$$dp_i = (2\pi \overline{e_{\nu_i}^2} \Delta\nu_i)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{v_i^2}{\overline{e_{\nu_i}^2} \Delta\nu_i}\right) dv_i. \quad \dots \quad (A3)$$

Applying the additive property of the "normal law," the probability of finding $v = \sum_i v_i$ in the range dv for the group of amplifiers is

$$dp = (2\pi \sum_i \overline{e_{\nu_i}^2} \Delta\nu_i)^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} \frac{v^2}{\sum_i \overline{e_{\nu_i}^2} \Delta\nu_i}\right) dv. \quad \dots \quad (A4)$$

In the limiting case, when $\Delta\nu_i \rightarrow d\nu$, and using eq. (15), it becomes

$$dp = (2\pi \overline{E^2})^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} \frac{v^2}{\overline{E^2}}\right) dv. \quad \dots \quad (A5)$$

Appendix II.

Often, when narrow band amplifiers are considered, the distribution of amplitude envelope of fluctuation voltage is of great interest. Let V denote the amplitude of fluctuation noise, then the probability of finding V in the range dV can be shown to be

$$dP = V/\overline{E^2} \cdot \exp\left(-\frac{1}{2} \cdot V^2/\overline{E^2}\right) dV. \quad \dots \quad (A6)$$

Thus, consider the figure where the instantaneous voltage v is plotted against the time t . The probability of finding the instantaneous voltage v in the range dv is equivalent to the fraction of time during which the voltage v will have the value v in the range dv , if the time considered is long compared to the inverse of the lowest frequency passed by the amplifier. Assuming that the distribution of v is given by eq. (17), then

the fraction of time d_1t , during which the voltage will have the value v in the range dv , is

$$d_1t \propto \exp(-\frac{1}{2} \cdot v^2/\overline{E^2}) dv. \quad (A7)$$

The fraction of time d_2t , during which the voltage will have the value $v + \Delta v$ in the range dv , is

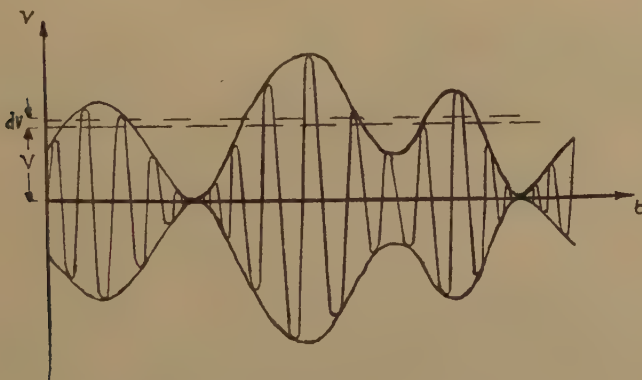
$$d_2t \propto \exp\left[-\frac{1}{2} \frac{(v + \Delta v)^2}{\overline{E^2}}\right] dv. \quad (A8)$$

The difference of time, $d\Delta t = d_2t - d_1t$, spent in passing through the maxima and the minima of the voltage curves in the range Δv , is

$$d\Delta t \propto -v/\overline{E^2} \cdot \exp(-\frac{1}{2} \cdot v^2/\overline{E^2}) \Delta v \cdot dv. \quad (A9)$$

When an amplifier of narrow band-pass (as compared to its mid-frequency) is considered, $d\Delta t$ is also the time during which the envelope has the value $V \equiv v$; so that for all possible values of v

$$\Delta t \propto -\Delta v \int_0^\infty V/\overline{E^2} \cdot \exp(-\frac{1}{2} V^2/\overline{E^2}) dV = -\Delta v, \quad (A10)$$



and the relative probability of finding the value V of the envelope in the range dV is

$$dP = \frac{d\Delta t}{\Delta t} = V/\overline{E^2} \cdot \exp(-\frac{1}{2} \cdot V^2/\overline{E^2}) dV. \quad (A6 a)$$

The probability that the envelope of fluctuation voltage will lie between zero and V is

$$P = 1/\overline{E^2} \int_0^V V \cdot \exp(-\frac{1}{2} \cdot V^2/\overline{E^2}) dV = [1 - \exp(-\frac{1}{2} V^2/\overline{E^2})], \quad (A11)$$

and the probability that the amplitude will exceed V is

$$Q = 1/\overline{E^2} \int_V^\infty V \cdot \exp(-\frac{1}{2} \cdot V^2/\overline{E^2}) dV = \exp(-\frac{1}{2} \cdot V^2/\overline{E^2}). \quad (A12)$$

The mean value of the amplitude is

$$\overline{V} = 1/\overline{E^2} \int_0^\infty V^2 \cdot \exp(-\frac{1}{2} V^2/\overline{E^2}) dV = \sqrt{\frac{\pi}{2}} \overline{E^2} \simeq 1.25 \sqrt{\overline{E^2}}. \quad (A13)$$

Bibliography.

- (1) Bernamont, J., *Ann. de Physique*, vii. p. 71 (1937); *Proc. Phys. Soc.* (London), xlix. (extra part) p. 138 (1937). Surdin, M., *Rev. Gen. d'Elec.* xlvii. p. 97 (1940).
- (2) Einstein, A., *Ann. der Physik*, xxxiii. p. 1275 (1910).
- (3) Landau and Lifschitz, 'Statistical Physics,' Clarendon Press, 1938.
- (4) Landon, V. A., *Proc. I. R. E.* xxix. p. 50 (1941); xxx. pp. 425 & 526 (1942).

LXXXIV. Notices respecting New Books.

Tables of Legendre Associated Functions. By ZAKI MURSI. [Pp. viii+283.] (Cairo, E. and R. Schindler, 1941.)

IF the ordinary Legendre function of the first kind is $P_m(x)$, the associated functions are $P_n^m(x) = (1-x^2)^{m/2} \frac{d^m}{dx^m} P_n(x)$, so that for integral n ,

$$P_n^m(x) = \frac{1}{2^n n!} (1-x^2)^{m/2} \frac{d^{n+m}}{dx^{n+m}} (x^2-1)^n.$$

Thus when $m=n$, $P_n^m(x)$ is $\frac{(2m)!}{2^m m!} (1-x^2)^{m/2}$ and is rational if m is even, or a rational multiple of $\sqrt{1-x^2}$ if m is odd; it does not exist for $m > n$. There are similar formulæ for Q_n and Q_n^m , the function of the second kind, non-regular at the origin, but these are not dealt with in these tables.

The values of P_1^1 and P_2^2 have been calculated from $x=0$ to 1 from this formula, and then P_m^m deduced from these by the recurrence relation $P_{m+2}^{m+2} = (2m+1)(2m+3)(1-x^2)P_m^m$. The values for $m < n$ were then obtained successively from the formula $P_m^{m-1} = (2m-1)xP_{m-1}^{m-1}$.

Dr. Mursi, who works in the department of pure mathematics at the Fouad I University, Cairo, has himself calculated all the entries, which cover $n=1$ to 10, $m \leq n$ and $x=0$ to 1 by intervals of 0.001, with an accuracy such that 8 decimals are retained except for $n=9$ and 10 and $m \geq 5$, where (the entries being large) 4 decimals (at least 9 significant figures) are given. In addition, Dr. Mursi made an independent calculation for 13 of the functions, S. A. M. Hassanein Effendi of 23 of them and M. N. Ghabbour Effendi of the remaining 19.

Since P_m^m were derived from P_1^1 and P_2^2 , any errors in earlier stages should be revealed by differencing P_{10}^m ; this was the first check applied. Then, Dr. L. J. Comrie differenced 6000 entries, and finally all entries have been differenced as far as the fourth order. The tables list the modified second differences. A check of a different kind was obtained by comparing the entries with those of Tallquist, published at Helsingfors in Finland in 1906. These tables are for $x=0$ by 0.01 to 1 and $n=1$ to 8.

The tables under review are well arranged; each page covers a range of 0.05 in x , the last entry on one page being repeated as the first of the next page, and the columns are arranged in the order $P_0^r, P_1^r, P_2^r \dots P_r^r, P_0^{r+1}$, and so on. At the end is a list of four errata, all affecting the 3rd or 4th figure of the second difference.

J. H. A.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

LXXXV. *Difficulties in Dirac's Representation Theory.*

By H. A. C. DOBBS*.

[Received March 11, 1943.]

(Concluded from p. 674.)

I must now go on to discuss those other principles connected with the assimilation postulate which lie at the root of the difficulties in Dirac's representation theory. These are :—

(1) *The general expansion theorem* covering the case of observables with a continuous range of eigen values. I have already said something on this topic in Part I. of this note, and I do not propose to say much more at this stage beyond laying emphasis on the obvious fact that the theorem is absolutely fundamental for the use of the eigen- ψ 's of an observable as the basic ψ 's of a representation. For all such representative systems require that an arbitrary ψ should be dependent linearly upon the basic states of the observables diagonal in the representation ; and the expansion theorem is required to ensure the required linear expansion for such an arbitrary ψ in terms of the simultaneous eigen states belonging to the set of diagonal observables. As we have already seen, the problem is a perfectly simple one in the case of a vector space with an enumerable number of dimensions—that is to say, in a representation for a discrete range of eigen values. In such cases the task of providing an expansion for an arbitrary ψ is simply the well-known problem of the reduction of a quadratic form to principal axes. But the continuous case is much more difficult, and Dirac admits that no formal or rigorous proof is available. He satisfies his conscience over this lack of vigour in his theory by making the assertion (page 37) : “The consistent development and physical interpretation of quantum mechanics require us to make the assumption that *only those Hermitian operators that satisfy the expansion theorem represent observables* and thus that *an arbitrary state is dependent on the eigen states of any observable.*” He then goes on to add (page 38) : “Hence if we know that a certain Hermitian operator represents some dynamical quantity which can be observed (for example if it represents the energy of some system) we may use the expansion theorem for this operator without fear of getting into error.” As far as I can see this argument appears to be the merest and most trivial tautology, its conclusion being simply a direct verbal consequence of the way in which the word “observable” is used. Obviously we shall feel just the same “fear of getting into error” as

* Communicated by Professor G. H. Hardy, F.R.S.

before, since we have no better guarantee against the existence of non-expandible observables after this principle has been asserted than we had before. This restrictive principle will of course (if it is *assumed* to be true) provide a basis for a representation theory of the kind Dirac wants for observables having a continuous range of eigen values. But Dirac has advanced no kind of rigorous argument or justification for it other than the remark on page 37: "there are usually physical grounds for telling when an application of the expansion theorem is permissible." What sort of things these "physical grounds" are he nowhere explains very clearly, and one is left in some doubt regarding the criterion for the application of the general expansion theorem.

On page 71 he again refers to "physical grounds" in connexion with the general representation theory, when he says: "From general physical grounds . . . one would expect the theory (of both discrete and continuous representations) to run on somewhat parallel lines in the two cases." One has a suspicion therefore that by the phrase "physical grounds" Dirac means just the requisite conditions for the valid application of the general representation theory, and consequently that his grounds for believing the expansion theorem to be applicable to the case of continuous observables are little more than the wish to assimilate the continuous and discrete representation theories. In fact, when in the course of his work he seems doubtful about the satisfactoriness of his theory, it is mostly in connexion with interaction problems involving radiation of very short wave-lengths. Here he admits that the wave-equations have, strictly speaking, no solution, because of the divergence of the series of Fourier components due to waves of very short length. For example in his discussion of field theory he points out that the classical expression for the total energy of a system of an atom interacting with a field of radiation $H_F = \frac{1}{8}\pi \int (E^2 + H^2) dX$ becomes, according to quantum theory,

$$H_F = \frac{1}{8}\pi \int (E^2 + H^2) dX - \frac{1}{2} \sum_a \hbar V_a. \quad (26) \quad [\text{XIII. (13)}]$$

Thus there is a term $-\frac{1}{2}\hbar V$ for each separate value of a , i. e., each separate degree of freedom in the system. But since the variable a is used to label the stationary states of the photons, and so consequently has a continuous range, the expression for the total energy of the interacting system is not a definite quantity at all.

There are other interaction problems where complete and rigorously accurate solutions are not obtainable by Dirac's methods, *e. g.*, the calculation of the total energy of a set of charged particles (such as electrons) interacting with a field of radiation. Here the Schrodinger wave-equation for the system can be written (non-relativistically) in the form:

$$i\hbar \frac{d\psi}{dt} = \{H_F + \sum_r H_r\} \psi, \quad \dots (27) \quad [\text{XIII. (50)}]$$

H_F being the energy due to the field alone and H_r being due to the

charged particle. Now we cannot give a finite expansion for ψ for the simple reason that $\frac{d\psi}{dt}$ is not a definite quantity at all, owing to divergent contributions arising from Fourier components with very short wavelengths. Dirac attributes this failure to the use of a point-charge picture of the electron (analogous to the classical model) which results in an infinite electric density in the neighbourhood. This may be the sort of consideration he has in mind when he talks of "physical grounds" as a criterion for applying the expansion theorem. If so, the criterion could equally well be stated in a purely analytic way in terms of the formal characteristics of the functions concerned. For what precludes a rigorous and exact solution is just the impossibility, *ex hypothesi*, of evaluating a Riemann definite integral over a discontinuous range, the discontinuity occurring in the neighbourhood of the point-electron. In such cases we must say, in accordance with Dirac's assumption, that the total energy of the system is not an observable; but I cannot see that he has shown us a *physical* criterion for deciding the applicability of the expansion theorem, which still reduces to the mathematical question of whether we can, in fact, actually perform the necessary integration. I am unable to see, therefore, that Dirac has given us any very cogent grounds for confidence in the general expansion theorem, as he has formulated it, unless we are prepared to acquiesce in a completely artificial restriction of the term "observable" to such operators as do *in fact* obey it. For on the one hand no rigorous, formal, considerations have been advanced in its favour; and on the other hand no satisfactory material criterion has been given for deciding when it may legitimately be applied. Thus we find an important class of the physical problems to which quantum theory ought to be applicable, but for which Dirac's technique offers no better solution than the classical electrodynamics. We are left uncertain whether this failure is ultimately attributable to a fundamental inadequacy of quantum concepts and technique; or whether it is due to an unsatisfactory formulation of the expansion theorem by Dirac (Dirac himself naturally inclines to the first alternative—*vide* his concluding remarks on pages 296 and 297, but it is one of my objects in writing this note to argue against this defeatist view). His own argument, based on the idea of a Fourier expansion for ψ , designed to make "the assumption appear plausible," is not very convincing because we know of many functions whose behaviour is such that we cannot give them a Fourier expansion (in terms of Riemann integrals). The integrations occurring in Dirac's theory are all Riemannian, and he nowhere makes any reference to the non-Riemannian developments initiated by Lebesgue and others. These extended methods have, however, enabled a Fourier analysis to be applied to a very much wider range of functions than was possible with Riemannian integration, and have greatly facilitated the determination of the coefficients in a Fourier expansion. Thus, I think, it is obvious that the plausibility of Dirac's expansion theorem could have been greatly strengthened by having recourse to the admission of

non-Riemannian methods of integration ; and, indeed, I am of the opinion that the further development of quantum theory along Dirac's lines could only proceed in this way. This procedure was not, however, open to Dirac himself for reasons connected with his assimilation postulate which I shall explain shortly.

(2) *The improper δ -function.* This is the second of the objectionable principles mentioned above as being partly responsible for the contradictory features in Dirac's theory. I have already shown in Part I. why Dirac found it necessary to introduce this concept, and I shall now examine the implications in more detail.

The first point to notice is the way in which Dirac introduces the δ -function. He gives its definition as

$$\int \delta(x) dx = 1, \quad \delta(x) = 0 \quad (x \neq 0). \quad \dots \quad (28) \quad [\text{IV. (2)}]$$

He adds (page 72) : "To get a picture of $\delta(x)$, take a function of the real variable x which vanishes everywhere except inside a small domain, of length ϵ say, surrounding the origin $x=0$, and which is so large inside this domain that its integral over this domain is unity. The exact shape of the function inside this domain does not matter provided there are no unnecessarily wild variations (for example provided the function is always of order ϵ^{-1}). Then, in the limit $\epsilon \rightarrow 0$, this function will go over into the δ -function."

Now this quotation makes it clear that the δ -function itself is to be conceived as vanishing everywhere except at the origin $x=0$. Consequently the notion of integration associated with the δ -function will require careful consideration. For clearly the content of the δ -function itself must be a set consisting of just one disconnected point at which the value of the function is infinite. This will preclude the function from being integrable in the Riemannian sense.

For (by definition) the Riemann definite integral is defined by

$$\int_a^b f(x) dx = \lim_{\Delta x \rightarrow 0} \sum_{x=a}^{x=b} (f(x_0) \Delta x) = \lim_{\Delta x \rightarrow 0} ((f(x_0) \Delta x + f(x_1) \Delta x + \dots + f(x_n) \Delta x). \quad \dots \quad (29)$$

Thus the idea of a Riemann definite integral is that of an enumerably infinite sum of vanishingly small *areas*, each of which is obtained by dividing up the projection of the curve of $f(x)$ on the x -axis between the limits of integration into a set of very small and equal segments, Δx in extent, and by multiplying these into a set of ordinates selected at an enumerably infinite set of points on $f(x)$. (It should be remembered that on the ordinary mathematical definitions a *continuous function* is completely determined by an enumerably infinite set of its values, *e. g.* by its rational values at all rational points.) It follows, therefore, that although the extent of Δx can be made as small as we please, yet it can never actually become strictly punctiform for two reasons. (1) If it did so then the terms such as $f(x_1) \Delta x$ would become identical with or

would reduce to terms of the form $f(x_1)$, which are of linear dimensions. Thus a $\int_a^b f(x)dx$ would cease to represent an area, *i. e.* would cease to be Riemannian. (2) An interval of finite extent can never be made to contract to an unextended point by any process of one-by-one subdivision or by any merely enumerable set of operations. If, therefore, we are to attach any significance to the notion of integration as applied to a function of this kind it must be in an extended non-Riemannian sense.

It is thus quite clear that the integral associated with the definition of the δ -function is not and could not by definition be Riemannian. But here we come up against the fundamental paradox in Dirac's formulation of the continuous representation theory. For he evidently thinks of the integral of the δ -function (and indeed of all the integrals of functions occurring in his book) as Riemannian. This is shown clearly by his way of developing the properties of the δ -function on pages 73-77 of 'Quantum Mechanics,' especially equations (6) and (15). Up to this point one feels that, even if Dirac's representation theory is not entirely satisfactory from a fully rigorous point of view, through its dependence upon an unproved general expansion theorem, yet it has considerable plausibility. But when he introduces the idea of the δ -function in order to secure his formal parallelism, then we begin to suspect the theory of becoming definitely invalid. Dirac would probably consider his opening remarks on improper functions at the beginning of Chapter IV a sufficient apology for their intrusion, and he tries there to treat their lack of rigour as a minor peccadillo. Thus, he says (page 72): "This means allowing that $(\xi'|)$ need not be a function of its variable ξ' according to the usual mathematical sense, which would require it to have a definite value for each value of its variable lying in a certain range, but may be something more general which we call an improper function of the variable ξ' . Such an improper function may be pictured as the limit of a sequence of ordinary functions, corresponding to the fact that a ψ which cannot be expressed in the form

$$\psi_x = \int \psi(\xi') d\xi'(\xi'|x), \quad (30)$$

with $(\xi'|)$ an ordinary function of ξ' , may be regarded as a limit of a sequence of ψ 's that can." (This point has some importance in connexion with his statement on page 79 *à propos* of basic ψ 's of a continuous observable: "The state represented by a basic ψ , $\psi(\xi')$, is not of a kind that can actually exist. . . . The state represented by $\psi(\xi')$ may, however, be regarded as limit of actual states and as such is a very useful theoretical abstraction.")

He also says: "We get over the difficulty (of finding a suitable expression for every ψ) by allowing the representative $(\xi'|)$ to involve infinities and singularities of a certain type chosen in just such a way as to make the expansion (30) always formally possible." And "The chief improper function which we shall have to deal with is the δ -function. . . ."

In fact, it is clear from all this that wherever the δ -function is needed for the expansion of any ψ , then the resulting representative is an improper function in Dirac's sense. Thus we can see that the impropriety of the δ -function will infect all functions expressed by means of representatives of a continuous observable. Consequently, we cannot employ Riemannian integrations in connexion with continuous representatives without fear of contradiction. Dirac himself does not take this view of the consequence of introducing such improper functions into the representation theory. I shall quote his statement on this point at some length because it gives a very clear picture of the kind of attitude which he tends to adopt throughout, and which is in my opinion one of the prime sources of the paradoxical elements to be found in his quantum theory. He says, on page 74: "The foregoing work shows how the expansion of a ψ as an integral in the form $\psi_x = \int \psi(\xi') d\xi'(\xi'|x)$ can be made of universal validity by the introduction of suitable improper functions. In this way we can get a foundation for the theory of representations in the case of continuous eigen values, corresponding to the foundation provided by the equation $\psi_x = \sum_r \psi_r(r|x)$ for the discrete case. Our definition and use of improper functions is not rigorous according to the standards of pure mathematics. It should be noticed, though, that an improper function can be given a rigorous meaning whenever it is a factor in an integrand. Now in the development of the theory, in every case where we have an improper function, it will be something which is to be used finally only in integrands. We could therefore rewrite the theory in a form in which the improper functions appear all through only in integrands and could then eliminate the improper functions altogether and make the theory rigorous. The use of improper functions is thus not really connected with any essential lack of rigour in the theory. It is rather a convenient notation, enabling us to express in a concise form certain fundamental formulas which we could, if necessary, rewrite in a rigorous form, but only in a cumbersome way in which the parallelism with the case of discrete eigen values would be obscured. . . ." Dirac's view of the δ -function is that it should be regarded merely as a convenient symbolic device—in fact, just an instance of what some writers (*e. g.*, Whitehead and Russell in 'Principia Mathematica') have called an "incomplete symbol." Now the valid use of an "incomplete symbol" depends upon:—(1) the possibility of giving a satisfactory definition of it in terms of certain other symbols which, in that context at any rate, can be treated as "complete symbols"; such a definition of the incomplete symbols in terms of "complete" ones must not, however, conflict with any other axioms used in the theory: (2) a proper specification of the permissible contexts in which the incomplete symbol can be significantly used.

These conditions imply that incomplete symbols are not to be constructed arbitrarily, nor applied inappropriately, or contradiction may result. We ought now to inquire whether Dirac's account of his

improper functions will meet these requirements. To do this we need only consider the δ -function itself, since all the other improper functions employed by him derive ultimately from this one in such a way that, if we cannot validly regard the δ -function as an incomplete symbol then, *à fortiori*, we shall not succeed in the case of the improper functions.

As regards the first point, the definition of the δ -function, this has already been considered and the conclusion was reached that it presupposes the notion of non-Riemannian integration. Now I hope to show when I come to discuss Dirac's use of the weight function $S'\Delta\xi'$ (standing for the number of discrete points lying in a small range $\Delta\xi'$) and its relations to the assimilation postulate, that this result of his definition of the δ -function is definitely inconsistent with other principles needed for the continuous representation theory. I shall therefore go on to the second question, *i. e.* the delimitation of contexts for the valid use of the δ -function regarded as an incomplete symbol. The fundamental equation for this purpose is

$$\int_{-\infty}^{\infty} f(x)\delta(x)dx=f(0). \quad . \quad . \quad . \quad (31) \quad [\text{IV. (3)}]$$

This equation, together with certain other important corollaries such as

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)dx=f(a), \quad . \quad . \quad . \quad (32) \quad [\text{IV (4)}]$$

Dirac deduces directly from his definition of the δ -function, and this is his basis for the statement regarding contexts already quoted which tells us: "It should be noticed though that an improper function can be given a rigorous meaning whenever it is a factor in an integrand."

Dirac attempts to establish

$$\int_{-\infty}^{\infty} f(x)\delta(x)dx=f(0)$$

by an elementary argument based on the picture of the δ -function associated with his definition on page 72. He uses this to argue that, since the sequence of functions $\phi_1(x), \phi_2(x) \dots$ (whose limit is $\delta(x)$)¹ vanishes everywhere outside a progressively diminishing existential domain, ϵ in extent, the left-hand side can only depend upon values of $f(x)$ very close to the origin $x=0$; consequently it will be valid to replace $f(x)$ by $f(0)$ without serious error, whereupon the theorem follows at once, for in the limit when $\epsilon \rightarrow 0$ the function $\phi_n(x) \rightarrow \delta(x)$. But we should note one important consequence of this reasoning. For it could only be valid if the function $f(x)$ were to conform to some kind of continuity standard—that is to say that, over any small interval ϵ in the domain of x (which we shall suppose to be the progressively diminishing domain of existence of the sequence of functions $\phi_1(x), \phi_2(x), \phi_3(x) \dots \phi_n(x)$, the limit of which is used to define $\delta(x)$), the variation in the magnitude of $f(x)$ must not exceed some number σ (say) which has a functional relationship to ϵ (it must be, for instance, of the same order of magnitude as ϵ). For if $f(x)$ failed to satisfy some such requirement there would be a

danger that the change in its value taken over the critical range ϵ round $x=0$ would *not* be negligible. In this case it would be invalid to replace $f(x)$ by $f(0)$ (*i. e.* to treat $f(x)$ as equivalent to $f(0)$) and the argument would fail. This would actually happen if $f(x)$ were to be discontinuous in the neighbourhood of $x=0$. Now Dirac has given us no kind of guarantee that he will only use the δ -function as an incomplete symbol in functional contexts which are up to the required continuity standard. Indeed, as he does not specify this standard, it would clearly be impossible for him to lay down any satisfactory general criterion to ensure this. What he does instead is to try and comfort us with the vague assurance, on page 74, that: "We shall confine our use of improper functions to such elementary equations that it will be obvious that the lack of rigour associated with them will not lead to a wrong result." Unfortunately he does not fulfil this undertaking very rigidly, and in the course of the work uses the δ -function in functional contexts, involving singularities, infinities and other irregularities, of a kind that clearly do not conform to the required continuity standards. For example, in the work in § 53 of the chapter on collision problems, Dirac develops an explicit expression for the polar representative in the p -representation of the outward-moving particles in a scattering system.

This representative, which Dirac writes $(P\omega)(\alpha'|1)$, involves in its explicit expression a term $\hbar^{3/2}(P\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0)/\{w'-w\}$. It is clear from the form of this term that the function will have a singularity or discontinuity at the point for which the factor $w'-w$ vanishes. This point, P' , is derivable from the equation $w^2/c^2 = m^2c^2 + P^2$, where P is the usual vector whose components are the momenta p_x , p_y and p_z canonically conjugate to the cartesian co-ordinates x , y and z . Now the singularity at P' in $(P\omega\chi\alpha'|1)$ is important for the determination of the angular distribution of the particles at infinity of the scattering system. In order to use it for this purpose it is necessary to give a meaning to the term

$$\hbar^{3/2}(P\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0)/\{w'-w\},$$

and this Dirac achieves by making, as he puts it (page 197): "... the convention that its integral with respect to P over a range that includes P' is the limit when $\epsilon \rightarrow 0$ of the *integral when the small domain from $P'-\epsilon$ to $P'+\epsilon$ is excluded from the range of integration.*" The words I have italicised make it clear that $(P\omega\chi\alpha'|1)$ is to be given a meaning by using a function (in this case a Riemannian integral) which is discontinuous over an infinitesimal range, 2ϵ in extent, in the neighbourhood of the point P' . But the representative $(P\omega\chi\alpha'|1)$ is also made up from a term involving the δ -function $\delta(w'-w)$, and in addition involves the representative of the associated ψ_0 , which in the ordinary p -representation is given by

$$(p\alpha|0) = \hbar^{3/2} \delta_{\alpha\alpha'} \delta(p-p^0). \quad . \quad . \quad . \quad . \quad . \quad (32)$$

Moreover, if it is desired to impose the condition that $(P\omega\chi\alpha'|1)$ is to represent only outward moving particles, we find (see page 200) that it

must contain a factor $1/\{w'-w-i\pi\delta(w'-w)\}$ equal to the improper differential coefficient of $\log(w'-w)$ when the pure imaginary values which vanish discontinuously at P' are taken into account. Here, then, we have a concrete instance of Dirac's association of the δ -function with a discontinuous expression, and we see clearly from this that Dirac does not hesitate upon the occasions when it suits him to employ the δ -function in connexion with other functions which do not conform to the continuity standards required by his treatment of it as an incomplete symbol.

Other instances of this will no doubt occur to the reader of 'Quantum Mechanics,' e. g. in his development of a limited quantum electrodynamics for systems of charged particles interacting with a field of radiation. Here his use of a point-charge model for the electron causes the wave-functions for the system to become infinite in value, owing to the divergence of the set of terms corresponding to components of the field of very short wave-length. Consequently, the wave-equation cannot be accurately solved in terms of Riemannian integrals so as to yield proper ψ representatives for the states corresponding to Fourier components of very short wave-length. For practical problems, where the wave-lengths considered are not too small, sufficiently satisfactory approximate results are obtainable by simply neglecting such infinities and irregularities due to the non-convergence of some components of the field functions. But these practical expedients in no way justify Dirac in his idea of applying a second quantization to the ψ 's occurring in the wave-equation of such systems; for this procedure would require the values of the ψ 's to be certain operators obeying a particular set of quantum conditions which, in the case of electrons, would require the use of the δ -function for their specification. Thus we see that Dirac's actual use of the δ -function is inconsistent with his attempt to treat it as an incomplete symbol having a definition in use that can be made fully rigorous. We must therefore conclude that this part of his representation theory is not admissible from a rigorous standpoint, and could only be made so by introducing non-Riemannian methods of integration. But, as I have already said, this course was not in fact open to him because it would have conflicted with certain requirements both of his assimilation postulate in general and of his " s " weight function in particular.

(3) *The " s " weight function.* It will be best to start the discussion of this third and last principle associated with the assimilation postulate by quoting Dirac's words on page 86 :—

"For a certain type of general theoretical investigation the use of a continuous range of eigen values in the representation is extremely inconvenient, and it becomes desirable, and is permissible, to replace the continuous range by a discrete set of points lying very close to one another over the whole range, and eventually to pass to the limit when the density of these points is everywhere infinite. This procedure is equivalent to the introduction of a certain weight function which tends to infinity in the limit. Let the number of discrete points in the small

domain ξ' to $\xi' + d\xi'$ (which may be either one-dimensional or many dimensional) be $s'd\xi'$, where $s' = s(\xi')$ is any function of ξ' which is everywhere large. Thus s' is the density of the discrete points. The general formula connecting a sum over the discrete points with an integral over the continuous range is now

$$\sum_{\xi} F(\xi') = \int F(\xi') s' d\xi', \quad \dots \quad (33) \quad [\text{IV. (41)}]$$

which shows that the discrete representation is equivalent to that continuous representation in which the weight function ' s ' has been introduced." The connexion of this principle with the assimilation postulate (page 71: "the possibility of regarding a continuous range of numbers as the limiting form of a discrete set whose density is increased indefinitely") is sufficiently obvious. Indeed, it is clear that the procedure implicit in this use of the " s " weight function is based entirely on the assimilation postulate, and the equation given above might be regarded as little more than a formal expression of this fact in the symbolism of the theory. It will be convenient, therefore, to complete this discussion, of the significance for Dirac's representation theory of the assimilation postulate (and the other fundamental equations associated with it), by devoting some attention to the fundamental equation (33), which for convenience I shall refer to as Dirac's weight theorem. One or two points about this expression strike us at once:—

(1) The connexion between the δ -symbol $\delta_{\xi'\xi''}$, the δ -function $\delta(\xi' - \xi'')$, and the weight function " s ."

For it can easily be seen that

$$\sum_{\xi' \text{ or } \xi''} \delta(\xi' - \xi'') s^{-1} = \delta_{\xi'\xi''}. \quad \dots \quad (34)$$

This relationship is used by Dirac at various points in his theory—one example may be found at the top of page 278—and it is important as showing exactly how the weight function " s " can be associated with the δ -function in bridging the gap in the formalism between the discrete and the continuous cases. I need not go further into this point, since I have already discussed at some length the role of the δ -function in this connexion. But it is well to note the peculiarly intimate relation between the notion of an improper δ -function and the " s " weight function.

(2) Dirac attaches much importance to the weight theorem as a means of passing from sums, over a discrete variable, to integrals over a continuous range, and *vice versa*. Thus he makes considerable use of it for getting rid of inconvenient sums and integrals in connexion with his theory of second quantization. But the form of the relation between the two in this weight theorem appears to be quite different from the sort to which we are accustomed in the theory of quadrature, and which is to be found for example in the well-known Euler-MacLaurin Theorem

$$\int_{x_0}^{x_1} u dx = \frac{h}{2} (u_0 + u_1) + \left[\phi(x) \right]_{x=x_0}^{x=x_1}, \quad \dots \quad (35)$$

which enables the value of a definite integral to be given in terms of a trapezoidal area and the sum of a series of differential coefficients of various orders. By comparison with such equations, which preserve a proper equivalence between the number of dimensions on both sides, Dirac's weight theorem looks rather queer, because there would, *prima facie*, appear to be a dimensional disparity between the left- and right-hand sides. Thus the left-hand, interpreted according to its usual significance, would be the sum of an enumerable set of the values of the function (*i. e.* if ξ' is one-dimensional it would be the set of ordinates $F(\xi')$, $F(\xi)$. . .). But the right-hand side, if it is regarded as Riemannian and integrated by parts, must represent a quantity which is dimensionally of the order of an area.

The function $s=s(\xi')$ is supposed to represent the number of discrete points between the values ξ' and $\xi'+d\xi'$. It will be convenient to put $\Delta\xi'$ to represent the range from ξ' to $\xi'+d\xi'$ in order to investigate the behaviour of $s(\xi')$ over this interval. We know that $s'=s(\xi')$ must, by definition, be large everywhere over this interval $\Delta\xi'$, and that it tends to infinity in the limit as $\Delta\xi'\rightarrow 0$. But I am now concerned with the actual manner in which $s'\rightarrow\infty$ as $\Delta\xi'\rightarrow 0$, for this is of great importance as determining the kind of manifold to which Dirac's weight theorem leads.

Since any $\Delta\xi'$ must be finite, the sequence of $\Delta\xi''$'s, whose limit is 0, must consist of a series of progressively diminishing domains of ξ' , each of finite extent. Let us arrange this sequence in order of diminishing magnitudes, and write it $\Delta\xi'_1, \Delta\xi'_2, \dots, \Delta\xi'_n, \dots$. We shall then have a series of s' 's corresponding to the series of $\Delta\xi'$'s, which we may write $s'_1, s'_2, \dots, s'_n, \dots$, where s'_1 is the value of $s(\xi')$ for the interval $\Delta\xi'_1$ (*i. e.* in the number of discrete points in $\Delta\xi'_1$), s'_2 is the corresponding value of $s(\xi')$ for the interval $\Delta\xi'_2$, and so on. The series of s' 's is of course just a progression of integers in ascending order of magnitude, and consequently must always remain *enumerable* (even when it becomes infinite), since the cardinal number of any progression can easily be shown to be the enumerably infinite number \aleph_0 (on Cantor's theory of transfinite numbers the smallest of all the infinite cardinals). But $s'd\xi'$ is to represent the number of points in the interval from ξ' to $\xi'+d\xi'$. Thus it follows that in the limit when $d\xi'\rightarrow 0$ the density of points per interval will be only an enumerably infinite one. This result, which is of course an inevitable consequence of assuming the assimilation postulate, shows that the employment of the " s " weight function would not be valid in any continuum in which a point-set can be homeomorphic with the set of real numbers. That is to say, neither the weight theorem nor the assimilation postulate would be admissible in a Cantor continuum, though they would be valid in manifolds having the sort of connexity exhibited by the "continuum" devised by Weyl in his book 'Das Kontinuum.' For in accordance with Weyl's principles of construction a "continuous" series is merely one which is everywhere dense or compact, so that the number of points in an extended interval need not

be more than \aleph_0 . But in a Cantor continuum the number of points in *all* extended intervals is the non-denumerable number 2^{\aleph_0} . Now a Weyl-like manifold can serve our purposes well enough when we are merely dealing with certain restricted portions of analysis (*e. g.*, such as the elementary forms of the infinitesimal calculus and such simple functions as the exponential and trigonometric). But this will not meet all the requirements of the modern theory of functions, since it demands a continuum with no gaps or irremovable irrationalities, and makes use of existence theorems (*e. g.*, the theorem of the upper bound) about real numbers which need not in general be explicitly definable by means of convergent series. But such theorems cannot be proved for a Weyl continuum (or any one similar to it) which comprises an infinity of "gaps."

We can now see that Dirac's general attitude to the question of "the continuum" is confused and ambivalent. For, on the one hand, in some places he draws a clear and explicit distinction between the "infinite enumerable" and the "infinite equal to the number of points on a line"—*vide* for instance page 71 and the footnote to page 82. But, on the other hand, he frequently finds it convenient to ignore this distinction—we have seen an example of this in his treatment of the eigen functions of the stationary states of the harmonic oscillator as transformation functions. Indeed, not only his transformation theory in general, but all those other parts of the representation theory which involve the assimilation postulate, require the treatment of a continuous stretch of points as a merely enumerable infinity.

Moreover, as we have seen, Dirac's account of the δ -function cannot be made satisfactory if the integrations involved are only Riemannian. But unfortunately none of the generalizations and extensions of the Riemannian methods, which were introduced by Lebesgue and his followers, are admissible in the sort of non-Cantorian manifold or "continuum" which is compatible with the use of the " s " weight function and the assimilation postulate. Dirac is thus placed in a very awkward dilemma by his continuous representation theory. For, on the one hand he cannot do without the assimilation postulate if he is to secure a close parallelism between the formalism of the discrete and continuous representation theories, and if he is to avoid the contradictions which must result in a Cantor continuum from his transformation theory. On the other hand the assumption of the assimilation postulate makes it impossible for him to introduce non-Riemannian methods of integration into his representation theory, since the sort of manifold to which the assimilation postulate inevitably leads will be incompatible with the generalized method of Lebesgue. Thus, only Riemann integrals will be permissible in a continuum to which the assimilation postulate and weight theorem is applicable, with the result that the δ -function and all the work dependent upon it must remain quite improper. Moreover, we can now see the justification for my previous statement that it is not

open to Dirac to use non-Riemannian extensions of the theory of integration to increase the plausibility of his expansion theorem.

We have now examined the assimilation postulate at some length and seen how it is related to the other basic principles of the representation theory, and why it must be considered theoretically unsatisfactory both from a rigorous theoretical point of view and also as a source of paradox and contradiction in the application of the representation theory. There is, however, one more point which must be touched upon before the matter can be finally disposed of. This arises out of the apparent similarity between the assimilation postulate and a principle which statisticians often invoke when employing a continuous distribution (such as the χ^2 distribution) to test the goodness of fit of a hypothesis relating to a discontinuous frequency distribution. It is not impossible that Dirac has something of this sort before his mind when he put forward the proposition which I have called the assimilation postulate as a basis for his continuous representation theory. But some readers may think that the resemblance between the assimilation postulate and the ideas underlying the mathematical theory of tests of goodness of fit is, *prima facie*, striking enough to call for some discussion, since the technique of fitting is generally regarded as a perfectly satisfactory part of the theory of statistics. Now some writers have certainly justified the use of the continuous χ^2 distribution, in tests of goodness of fit for a discontinuous frequency distribution, by asserting that a continuous distribution is *in fact* the limit to which a true discontinuous frequency distribution tends as the size of the sample is increased indefinitely.

This is said to be the rationale behind such devices as Yates's correction for continuity—a useful adjustment which is frequently used in connexion with discontinuous frequency distributions derived from small samples.

Yates's correction is made by entering the table of χ^2 not with the values calculated from the actually observed frequencies but with the values it would have if the observed frequencies had been less extreme than they really were by half a unit. Thus in a 2×2 frequency table the two larger divisions would be reduced and the two smaller increased by a half-unit. The idea behind this adjustment is of course to make the values of the frequency distribution actually observed approximate more closely to that ideally "true" discontinuous distribution which, in accordance with the principle mentioned above, statisticians suppose approaches the continuous distribution in the limit as the size of the sample is indefinitely increased. There may therefore be a temptation to say that in these cases it has proved possible to make a successful appeal to a principle which is essentially the same as the assimilation postulate. It will then be argued that if this procedure is permissible in one branch of applied mathematics it must be equally so in other cases, of which Dirac's continuous representation theory for quantum mechanics would be one. But I think that closer scrutiny will reveal that in fact there is no such analogy. To see that this is so we must try and discover what it is we are trying to do when a continuous χ^2

distribution is used for tests for goodness of fit with a discontinuous frequency table. For this purpose the value of χ^2 may be taken as $\Sigma\left(\frac{x^2}{m}\right)$, where x is the divergence from the number expected, antecedently, in any frequency class on the basis of the hypothesis. (The summation extends of course over all classes.) It is clear that the estimate of χ^2 will depend partly upon calculations in terms of the parameters attributed to the population (on the basis of the hypothesis to be tested), and partly upon its sampling variance. Thus it is clear that the estimates of χ^2 for a given population may vary to a considerable extent, and some of these estimates can definitely be said to be more "erroneous" than others. Now the fundamental condition that governs all tests for goodness of fit is that "errors of fitting" should not be comparable to the errors of random sampling; and this is equivalent to the requirement that the statistics which are used as an estimate of the parameters of the population should be efficient in the sense that they must make use of all available information.

When the statistician has satisfied himself that his statistics satisfy this essential requirement, he may go on to calculate χ^2 for the hypothesis. He will then be in a position to use the functional relationship which has been established between the probability P and the value of χ (whereby P is a continuous function of χ^2) to estimate the antecedent probability of obtaining this value of χ^2 for a frequency distribution of that kind, *i. e.* with that particular number of degrees of freedom. This value of P will enable him to judge whether or not his hypothesis is open to suspicion on the grounds of too great a divergence between the results of observation and expectation.

For example, if the value of P is found to be between 0.9 and 0.1, it is generally held that there is no strong reason to doubt the hypothesis on these grounds. But values of P greater than 0.9 or less than 0.1 are generally regarded with some suspicion. We can now see that, in this procedure for testing the worth of an hypothesis, at no stage do we want an *exact* estimate of the antecedent probability P . All we are concerned with is to make sure that the probability lies *within reasonable limits*. Thus a limited degree of inaccuracy, due to applying a continuous distribution to test the dispersion of a discontinuous frequency table, need not be practically objectionable, having regard to the purpose for which the test is designed.

Moreover, since the value of the actual index of dispersion calculated from a discontinuous frequency distribution must always be rational and of the form $\Sigma\left(\frac{x^2}{m}\right)$, we could obviously express the inaccuracy of using a continuous χ^2 distribution to calculate P in tests for goodness of fit by means of an expression $\epsilon(n)$, which was a function of the number of units in the sample (and which steadily *decreases* as $n \rightarrow \infty$); just as we can express the inaccuracy of summing a convergent infinite series

to a finite number of terms by a function involving the number of terms taken. But there are two points to notice here :—

- (1) This procedure is only satisfactory when the index of dispersion is being calculated for the "true" discontinuous distribution or for one very closely approximating to it. This means in practice that we must be very careful to guard against sampling irregularities produced by small numbers which may lead to very significant divergences (from the "true" discontinuous distribution obtainable from *large* samples) on the part of the actual distribution from which the index of dispersion is calculated. In practice this safeguard can generally be secured by seeing that in no division shall the expected number be less than five. In this way it is easy to ensure that the number of possible sets of observation shall be large, each occurring with only a small frequency. But exceptional cases will arise in the case of small numbers, or where a frequency distribution has only one degree of freedom, and these require special treatment. It is here that devices such as Yates's correction come in useful for the purpose of re-arranging the actually obtained results so as to smooth out irregularities due to small numbers, thus obtaining a distribution approximating more closely to the true discontinuous distribution.
- (2) The fact that the inaccuracy, inseparable from applying a continuous distribution to test a discrete frequency table, can be expressed as a steadily decreasing function of the number of units of the sample (and thus can be eliminated in the limit $n \rightarrow \infty$), depends entirely upon the sense of "continuous" appropriate to a *function* (and not to a *series* or *aggregate*); a continuous function being completely determined by an enumerable set of conditions, *e. g.* by its values at all rational points in its domain.

When the "continuity" of the χ^2 distribution is contrasted with the discreteness of a frequency distribution, what is usually really meant is this :

- (1) χ^2 can have any value between 0 and ∞ ;
- (2) P can be shown to be a continuous function of χ^2 , so that, for any value of χ^2 between 0 and ∞ , P has a particular value between 1 and 0.

Now this is not the sense of "continuous" or "continuity" in which an aggregate, series, or point-set such as a Euclidean line can be said to be a "continuous" range or to possess "continuity." This latter sort of continuity (which was first thoroughly investigated by Georg Cantor) is characteristic of the real number series and continuum required by the extensions of modern analysis, and depends, as we have already

observed above, on the notion of a set or class having a non-denumerable number of terms.

But it is with "continuity" and the "continuous" in this latter sense (appropriate to *sets* and not to *functions*) with which the assimilation postulate is concerned, as can easily be seen by referring to the words in which it is expressed.

We can now see why the principles, to which statisticians successfully appeal to justify the technique of tests of fitting by means of the "continuous" χ^2 distribution, give no support to Dirac's assimilation postulate; and to suggest that they may do so is to be guilty of a gross confusion between the fundamentally distinct senses of "continuity" and "continuous," which apply respectively to functions and to series or sets (actually mathematicians have not always distinguished these two senses as clearly and consistently as they should have).

The conclusion which thus finally emerges from this discussion of Dirac's assimilation postulate is this: if we assume the postulate to be true, in the form in which it has been stated, we must reject the notion of a Cantor continuum and conclude that Dirac's continuous representation theory is invalid (based as it is upon the notion of an improper δ -function which must be invalid in a Weyl-like continuum). But in this event, since the actual results of observation must form a *discrete* schedule of numbers, we should conclude that only those equations which refer to discrete representations can be validly applied to the real world as revealed by experiment. This conclusion would derive some support from his statement on page 79: "The state represented (in a representation for continuous observables) by a basic ψ , $\psi(\xi')$, is not of a kind that can actually exist. If an observable such as ξ' with a continuous range of eigen values is measured for any actual state, the result must be distributed over a finite range according to some definite probability law, which range may be made as small as we please but cannot be contracted to a single point. The state represented by $\psi(\xi')$ may, however, be regarded as a limit of actual states and as such it is a very useful theoretical abstraction." If these conclusions be accepted, however, it will only be permissible to work in quantum theory with finite difference equations in place of the differential equations of ordinary mathematical physics. That is to say, the calculus of finite differences will have to replace the ordinary infinitesimal calculus as the instrument for unravelling and co-ordinating the quantitative connexions and functional dependencies between observational phenomena. But unfortunately, the calculus of finite differences is a much less powerful tool for handling quantitative relationships than the infinitesimal calculus; and the methods and theorems of the latter could only be applied through the use of interpolations depending upon those formulæ which enable differential coefficients to be expressed in terms of a series of advancing differences. This would forbid us to express the relationship between discrete and continuous representations by means of transformations or transformation functions; we should have to do it rather by means of

interpolation equations, neglecting the higher order differences falling beyond the limits prescribed by Heisenberg's uncertainty principle. But this conclusion must have serious consequences for Dirac's formulation of the principles of quantum theory, and indeed it must destroy some of the most attractive features associated with his account. It is not possible within the limits of a note of this kind to exhibit these effects in full detail; but quite a good instance of the sort of thing I have in mind is to be found in his treatment of second quantization and its application to the theory of Einstein-Bose assemblies. Dirac makes use of this to reconcile the wave and corpuscular pictures of light by showing that any large Einstein-Bose assembly can be treated on the principles of second quantization as being dynamically equivalent to a set of harmonic oscillators. He defines the general procedure on page 234 as follows:—"This (*i. e.* the procedure of second quantization) consists in assuming that the wave-function ($q^{(a)}|$) describing the single system is not a numerical function of the parameter $q^{(a)}$, but is an operator for each $q^{(a)}$ satisfying the quantum conditions (15). It then goes over into ξ_a , the form of its wave-function and transformation law remaining unaltered." This principle asserts in effect that for particular purposes wave-functions referring to a continuous representation can be treated as equivalent to certain operators ξ_a, ξ_b , defined so as to conform to the quantum conditions:—

$$\left. \begin{aligned} \xi_a \xi_b - \xi_b \xi_a &= 0, \\ \bar{\xi}_a \bar{\xi}_b - \bar{\xi}_b \bar{\xi}_a &= 0, \\ \xi_a \bar{\xi}_b - \bar{\xi}_b \xi_a &= \delta_{ab}. \end{aligned} \right\} \quad \dots \quad (36) \quad [\text{XI. (15)}]$$

Now it is clear from the occurrence of the δ -symbol in these conditions that the ξ_a variables must operate on a discrete representation. Indeed, this would be sufficiently obvious from a consideration of their definition, which is given by

$$\xi_a = (n_a + 1)^{\frac{1}{2}} e^{-i\omega_a} = e^{-i\omega_a} n_a^{\frac{1}{2}}, \quad \dots \quad (37) \quad [\text{XI. (13)}]$$

for n_a is a variable representing the number of systems in the state $q^{(a)}$, so that it is clear that the eigen values of the n 's can only be integers. When defining the ξ -operators we can treat the n 's as observables (as we are entitled to do on Dirac's theory) and choose a representation in which they are diagonal, which must then be a discrete one. We can then see that the procedure of second quantization must, according to Dirac, involve the possibility of interpreting one and the same expression in terms both of wave-functions (referring to a continuous representation) and dynamical operators relating to a discrete representation. This of course implies the treatment of a continuous representation as homeomorphic with a discrete one—a procedure which is only justifiable by an assumption of the assimilation postulate. Actually, it would be technically impossible to apply a second quantization of the kind Dirac wants to such problems without the use of the assimilation postulate and his weight theorem (which between them enable the inconvenient

passages from sums to integrals to be made). How this is actually accomplished in a concrete case can be seen clearly in Dirac's application of a second quantization, on pages 238–241, to a large Einstein–Bose assembly. He is there investigating the interaction of a large number of photons with an atom, treating the photons as the systems of the assembly and the atom as the perturber. He takes the co-ordinate variables q to be constants of the motion, using the $q^{(a)}$'s to label the stationary states of the photons, and the n_a 's to represent the number of photons in the corresponding stationary state $q^{(a)}$. Then the Hamiltonian of the system can be written $H = H_P + H_R + H_Q$, where H_P is the energy of the atom alone, H_R is the total proper energy of the radiation, and is equal to $\sum_a n_a \hbar V_a$, and H_Q is the total interaction energy. This must be of the form

$$\begin{aligned} H_Q &= \sum_a \{ \bar{\xi}_a V_a + \bar{V}_a \xi_a \} + \sum_{ab} \bar{\xi}_a V_{ab} \bar{\xi}_b \\ &= \sum_a \{ V_a n_a^{\dagger} e^{i w_a} + \bar{V}_a (n_a + 1)^{\dagger} e^{-i w_a} \} \\ &\quad + \sum_{ab} V_{ab} n_a^{\dagger} e^{i w_a} (n_b + 1)^{\dagger} e^{-i w_b}, \quad \quad (38) \end{aligned}$$

where the V_{ab} 's are matrix elements of the representation, which are not pure numbers but consist of certain functions of the dynamical variables of the atomic system, and the V_a 's represent the interaction energy of the Q 's with the atomic system. But in this representation the q 's are constants of the motion, and are taken to be a complete set of commuting observables, representing the momentum and polarization. Now the n_a 's in the above equations are just the number of photons in the corresponding states $q^{(a)}$, so that the parameter a in both n_a and $q^{(a)}$ must be regarded as having a *continuous* range. This means that our expression for H_Q cannot be an accurate one for the interaction energy until sums are replaced by integrals, and to do this properly would involve adopting a particle picture having continuously many degrees of freedom.

Dirac gets out of the difficulty in the usual way by using the idea that a continuous range of points on a line is equivalent to the limit of a discrete set of an infinite density. Thus he introduces a weight function s_a equal to the number of *states* per unity of K -space in the neighbourhood of the value K_a , where s_a is a large but arbitrary function of K_a . (Incidentally to each value of the suffix a in n_a there corresponds a value of K , K_a , which represents the momentum and polarization.) Now besides the number of *states* about the value K_a we can introduce the idea of the number of *photons* (in a particular state of polarization) about the value K_a , and this can be written $\eta_a = n_a s_a$, provided n_a is a sufficiently continuous function of the schedule of states to which it refers.

Dirac now writes the whole expression in terms of representatives, and uses the matrix $(a|V|b)$ to represent the interaction energy, putting $(a|V|b) = V_{ab}(s_a s_b)^{\dagger}$; thus making use of the function s as a weight

function. Then, using this weighted representation, the expression for the interaction is

$$H_Q = \sum_a \{ (a|V|0) \eta_a^{\dagger} e^{i w_a} + (0|V|b) \eta_a^{\dagger} e^{i w_a} \} s_a^{-1} \\ + \sum_{ab} (a|V|b) \eta_a^{\dagger} \eta_b^{\dagger} e^{i(w_a - w_b)} s_a s_b^{-1}, \quad (39) \quad [\text{XI. (34)}]$$

and in the limit when s becomes infinite this expression passes over to

$$H_Q = \sum \int \{ (a|V|0) \eta_a^{\dagger} e^{i w_a} + (0|V|a) \eta_a^{\dagger} e^{-i w_a} \} dK_a \\ + \sum \int \int (a|V|b) \eta_a^{\dagger} \eta_b^{\dagger} e^{i(w_a - w_b)} dK_a dK_b, \quad (40) \quad [\text{XI. (35)}]$$

where the sums are over the discrete polarization variable. The whole of this procedure clearly depends for its validity upon the idea that a discrete set of points can be assimilated in the limit to a continuous set.

Thus we can see that Dirac's application of the procedure of second quantization, to problems involving the interaction of a set of particles with a field of radiation, essentially depends upon ideas derived from the assimilation postulate and its corollary the weight theorem. These principles make it possible to conceive of valid *transformations* between a continuous representation, with an enumerable number of degrees of freedom, and a discrete representation, having continuously many degrees of freedom; and the difficulties inherent in such transformations (which manifest themselves by the appearance of awkward sums or integrals) are avoided by Dirac's technique of the s weight function, which in its turn depends on the assimilation postulate.

It is clear that if we reject the assimilation postulate, and thus also the continuous representation theory founded upon it, we shall not only rob Dirac's exposition of quantum theory of much of the attraction it derives from its formal completeness and the elegant parallelism between continuous and discrete representations, but we shall undermine the very fabric of his general treatment of fundamental quantum principles. For this treatment involves the use of continuous representations as a basis for applying the infinitesimal calculus (and all the other analytical techniques applicable to continuous variables) to observables. In short, the rejection of the assimilation postulate will entail the loss of many of those merits which are characteristic of Dirac's presentation of quantum theory, and to which I referred at the beginning of this note. To avoid this result certain drastic remedies are necessary: in particular certain fundamental innovations and modifications will have to be made in the basic ideas which underlie the general representation theory. I shall close this note (already considerably longer than was originally intended) with a very brief indication of the directions in which such changes must lie.

First of all, in order to validate the representation theory for observables with a continuous range of eigen values, it would be necessary to introduce a form of non-Riemannian integration which would be compatible with an improper function such as the δ -function. Moreover, since the non-Riemannian extensions in the theory of integration result

in a considerable addition to the range of functions with a Fourier expansion, their introduction into Dirac's theory would considerably strengthen the credibility of his argument for the general expansion theorem.

Secondly, it would be necessary to admit that the idea of a diagonal representation for continuous observables involves the notion of a continuum with a non-denumerable number of co-ordinate axes or dimensions (*vide* Dirac's own remarks on this score at page 71). From this conception to the idea of a dynamical system with continuously many *degrees of freedom* is a step of but little theoretical consequence. Thus, if we accept the practicability of using continuous representations at all, we must be equally prepared to accept the correlative notion of dynamical systems with 2^{\aleph_0} degrees of freedom. This seems to be a new and repugnant idea to Dirac, although it is already implicit in his whole conception of a continuous representation, as we can see from his remark on page 239: "To do this accurately would not be very easy since it would mean dealing according to quantum mechanics with a dynamical system with continuously many degrees of freedom which would require a new notation and a new mathematical technique." (My object in quoting this passage is not to suggest that Dirac over-estimates the difficulties in the way of a rigorous development of these ideas; I merely wish to emphasize that the difficulties he mentions have actually been present all along and are implicit in his own continuous representation theory.)

Thirdly, the relation between the concepts of the discrete and the continuous should be approached from an entirely different point of view. In my opinion this new approach should be based on a recognition of the fact that there are two distinct but fundamental dichotomies underlying modern mechanics or theories of matter. These may be classified in this way:

- (1) the wave-particle dichotomy,
- (2) the discrete-continuous dichotomy.

The connexion is on these lines.

The quantities which result from taking observations in actual experiments are really only amenable to *discrete* representations. But owing to the tendency to refer discontinuous phenomenal manifestations to spatio-temporally persistent entities which are conceived as forming the enduring substrata of such manifestations (this is the essential feature of the particle picture), we reach *continuous* representations which are consequently applied (inappropriately in a sense) to the set of discrete observations. But the results of such calculations and operations, which are related to continuous representations, are not in reality applicable to the world of observation. Consequently, we find an indeterminacy of application associated with our differential equations. This indeterminacy actually proceeds from the converse of the process of interpolation (whereby relations between an essentially discontinuous schedule of

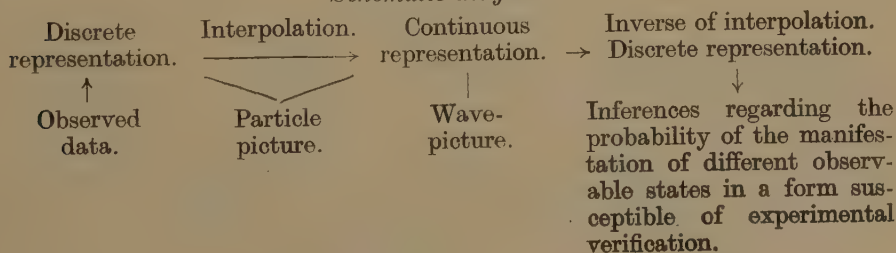
measured quantities are incorporated in differential equations relating to continuous representations).

We may thus conceive of the particle picture as being the idea of an interpolation nexus relating schedules of discontinuous quantities to a continuous representation. The wave-picture comes in when quantum theory substitutes the motion of a wave-packet for that of a particle. Now the particle picture envisages the possibility of a (theoretically) complete simultaneous determination of the values of all the dynamical variables of the system. Such a conception is not admissible in quantum principles. For this kind of completely determinate state (with which classical dynamical theories deal) quantum theory substitutes the idea of a wave-function representing a state for which the co-ordinates and momenta have approximate numerical values throughout all time. Such a state is of the form of a wave-packet, its equation of motion being given by a curve identical with the trajectory of the particle in the classical mechanics for the corresponding state in the particle picture. Indeed, as Dirac is able to show in his account of quantum principles, we can regard the classical equations of motion as the limit to which the quantum ones tend as the magnitude of the factor \hbar , occurring in Heisenberg's uncertainty principle, becomes vanishingly small in comparison with the other numerical quantities in the system.

Of course it is easy to show that the degree of determinacy with which values can be simultaneously assigned to dynamical variables can never remain at the limits prescribed by Heisenberg's uncertainty formula $\Delta q' \Delta p' = \hbar$. For this relation only applies to a wave-packet, in a maximal degree of concentration, from which it must diverge in the course of time owing to spreading. Thus, for the purpose of handling the discrete quantities of experiment and observation by the differential calculus (and other analytical techniques applicable to the continuous), quantum theory adopts a continuous representation, but substitutes a *wave-packet state* for the completely *determinate state* of the classical particle-dynamics. The wave-equations then enable us to predict the behaviour of these wave-packets and to use this knowledge to determine the probability of observing any particular state by actual experimental methods. This latter process of inference consists of *extrapolating* from the continuous to a discrete representation again, and forms the basis of all applications of quantum theory.

We can now represent the whole procedure schematically in the following way :—

Schematic diagram.



This diagram illustrates the crucial importance of interpolations ; and we see clearly that in order to develop quantum theory we must have a more accurate and powerful interpolation technique of a kind in harmony with quantum principles. This must enable us to pass from expressions relating to discrete representations (involving sums) to the corresponding expressions referring to continuous representations (involving integrals which may be *non-Riemannian*). Such a technique of quantum interpolation should enable us to substitute accurate quantitative relations between the finite differences of the discrete representations and the differentials of the continuous representation ; and (instead of Dirac's inaccurate methods based on the "s" function) would embody a proper and valid weight theorem giving the correct relations between all representatives (including sums and integrals) in the two different kinds of representation. The use of continuous representations would then be permissible whenever the required interpolation could (in principle) be performed. And the employment of non-Riemannian integrals and accurate interpolation equations in connexion with such representations should enable some of the difficulties (associated with functions which prove difficult to handle by the ordinary classical Fourier analysis and which limit the applicability and usefulness of Dirac's theory) to be avoided. *À propos* of these he says, on pages 296 and 297 :

"The limitations in the applicability of quantum electro-dynamics thus correspond precisely to those of classical electro-dynamics. The amendments required in classical theory, in order to make it apply accurately to the elementary charged particles, are thus not provided by the passage to the quantum theory, that is by the taking into consideration of the distances accompanying measurements. It seems that some essentially new physical ideas are here needed."

My object in writing this note has partly been to suggest a possible direction which such new ideas must take, and I shall therefore close with this quotation from Dirac.

LXXXVI. *Some Spectrophotometric Investigations on Iron Oxide-containing Soda-Lime-Silica Glasses.—Part I. The Influence of Concentration and Ferric-Ferrous Dissociation**.

By T. H. WANG, M.Sc., D.I.C., and Prof. W. E. S. TURNER, F.R.S. †.

[Received March 24, 1943.]

ABSTRACT.

The series of soda-lime-silica glasses used previously by Densem and Turner in their study of the equilibrium between ferric and ferrous

* A Paper read at the Sheffield Meeting of the Society of Glass Technology, Nov. 18th, 1942 and published also in the *Journal* of that Society.

† Communicated by Prof. W. E. S. Turner, F.R.S.

oxides in glasses has now been investigated spectrophotometrically in respect of the visual region. The parent glass had the composition 75 per cent. SiO_2 , 10 CaO , 15 Na_2O , and iron oxide, from 0.002 to 12.5 per cent. expressed as Fe_2O_3 , was progressively introduced. At concentrations below 0.6 per cent. of iron oxide a linear relation between absorption and degree of ferric oxide dissociation held approximately true within the range of dissociation of about 13–35 per cent. Extrapolation showed that the ferric iron at low concentrations contributes a greenish-yellow colour, whilst ferrous iron bestows a blue colour on the glass. At high concentrations the ferric iron exercised high absorption, particularly in the blue-violet end, thus resulting in a brown or pale brown colour. This high absorption component is sensitive to heat-treatment.

The results provide evidence that in attempts to formulate a theory of the decolorizing of glass, deductions based on glasses containing other than small amounts of iron oxide are inadmissible.

I. GENERAL INTRODUCTION.

In a previous communication one of the present authors * has described an apparatus for visual spectrophotometry which has been built up recently in this Department. It has proved especially suitable for the study of low absorptions in glasses caused by the presence of traces of colouring agents, and hence also for the investigation of the process of decolorizing glasses. Hitherto, most of our knowledge about the latter subject has been derived either from spectrophotometric studies of colouring agents at comparatively high concentrations, or, when the concentrations of such agents were of the order usually occurring in practice, by resorting to visual inspection or colour-matching. When the former process was applied, the question naturally arose as to how far the conclusions thus deduced could be applied directly to the low concentrations appropriate to the decolorizing range; and when visual inspection and colour-matching were employed, the results observed were only qualitative, and were insufficient to provide definite information on the effects of various constituents present. The apparatus referred to has provided the means for tackling the problem on a proper quantitative basis.

In most cases the necessity for the decolorization of glass arises from the unavoidable presence of small amounts of iron oxide. As a starting point we have first studied glasses containing different amounts of iron oxide, to determine how far the behaviour of iron at concentrations of, say, 1 per cent., differs from that when present to the extent of 0.1 per cent. and less. We hope in a later communication to refer to the absorption of soda-lime-silica glasses containing iron oxide in conjunction with constituents normally added in decolorizing processes.

* T. H. Wang, *Phil. Mag.* xxxiv. p. 684 (1943).

II. THE EXPERIMENTAL PROCEDURE.

1. *The Glasses Tested.*

The glasses studied comprised the parent, of composition 75 per cent. SiO_2 , 10 CaO , and 15 Na_2O , and derived glasses containing iron oxide ranging from 0.002 to 12.5 per cent. expressed as Fe_2O_3 . This series was prepared by N. E. Densem and W. E. S. Turner† with great care in the year 1935–36. Most of the specimens were in rod form. After such lapse

TABLE I.

Variation of Colour with Iron-Oxide Concentration in the Soda–Lime–Silica Glasses (75 SiO_2 , 10 CaO , 15 $\text{Na}_2\text{O} + x \text{Fe}_2\text{O}_3$). Melted at 1400° .

Total Iron Expressed as Fe_2O_3 , %.	FeO Total Iron (as Fe_2O_3), %.	Colour through 10 cm. length.
0.002	—	Slight yellow tinge.
0.022	27.2	" " "
0.033	27.2	Colourless.
0.042	35.6	Very pale blue-green.
0.053	34.0	Pale blue-green.
0.062	30.5	" "
0.077	30.2	Blue-green.
0.102*	27.0	"
0.19*	20.0	Green.
0.26*	17.2	Uniform increase in density of colour from 0.20 to 0.90%.
0.36*	16.0	
0.61*	14.8	
0.76	14.0	
0.98*	13.3	Olive green.
1.9*	11.4	Deep olive green.
2.8*	10.3	"Black."
4.67	9.2	"
6.1*	8.5	"
6.7	8.9	"

In the above table only those data which will be used in the following description are quoted. In the first column quoting the total iron, those values marked with an asterisk were from new determinations made on the specimens used in the present experiments.

of time it is possible that some slow change in the absorption characteristics of the glasses may have taken place, but no such change could have been due to solarization, since they had been stored carefully in the dark.

In the preparation of the glasses the raw materials used were quenched quartz obtained from selected rock crystal, A.R. pure anhydrous sodium carbonate, and A.R. brand precipitated chalk; and, prior to melting,

† J. Soc. Glass Tech., Trans. xxii. p. 372 (1938).

the well-mixed batch was sintered for 6 hrs. at 730° in a silica crucible in an electric muffle furnace, in order to reduce to the lowest limits any subsequent slight solution of platinum in the glass. The melting of the sintered batch was carried out at 1400° , in a platinum crucible in an electric furnace, the constant melting time of 22 hours being adopted. Specimens were drawn during the last hour of melting. The chemical analyses and colours of the glasses, as stated by Densem and Turner, are recorded in Table I.

2. *The Spectrophotometry.*

The apparatus has already been described*, and it will suffice to mention those points only which are essential to the present investigation.

Throughout the measurements the slit of the spectrometer was kept at the same width, corresponding to a wave-length range of approximately 20, 60, and 90 Å. at wave-lengths 4500, 5893, and 6500 Å., respectively. The photometer field, of the juxtaposed spectra type, was opened to the same extent, so that the effective slit widths were slightly greater than the wave-length ranges of the slit. No attempt was, however, made to apply the slit-width correction to the observed data.

The wave-length calibration of the instrument was checked by means of a sodium lamp and a hydrogen discharge tube. The photometer scale was checked by means of fixed-angle rotating sectors. Throughout the measurements the substitution method was used to determine the transmittance.

Particular mention should be made that the instrument was not subjected to any measurable systematic error arising from the presence of stray light at the ends of the visible spectrum. Repeated check determinations obtained in various ways, as, for example, by the insertion of a filter, further verified this point.

In calculating the extinction coefficient of the glasses, the reflexion loss was determined from the Fresnel's reflexion factor and deducted from the transmittance measurements by assuming the index of refraction of the glass specimens to be 1.52 throughout. Since the specimens used were fairly long, for low absorptions a slight change of the refractive index due to the introduction of iron oxide would not appreciably affect the value of the extinction coefficient calculated. At high concentrations the absorption of the glass became so intense that even if the specimen used were thin, the relative error introduced was still quite inappreciable, as can be deduced from the results described below.

The preparation of the specimens followed precisely the procedure described in the preceding paper†. With one or two exceptions, when the circumstances warranted it, the ends of specimens were polished, and when slabs were used, polished plane parallel.

* T. H. Wang, *loc. cit.*

† T. H. Wang, *loc. cit.*

A few comments may be made about the measurements at the two ends of the visible spectrum. Due to the use of the broad uniform source, which is a particular feature of the apparatus designed for the investigation, the brightness was considerably lower than that obtainable by a Pointolite lamp as used in ordinary practice. As the result, the spectrophotometric measurements at the two ends of the spectrum could only be made when the absorption in these regions was low. With the iron oxide-containing glasses, however, this condition was, in general, not fulfilled. Consequently, the determination at the two ends cannot be pushed very far without risk. To remedy this, a second device was used in conjunction with the conventional method. In this device a Pointolite lamp was employed as the source. A lens was placed in front of the lamp to obtain a bigger apparent source. In these circumstances, due to undue refractions of the rays when passing through the specimen, only part of the essential rays entered the aperture of the instrument. Thus, the photometric setting could not give the true value. Nevertheless, if there were no changes in refractive index with respect to change of wave-lengths, the condition of the rays within the specimen would remain the same for all wave-lengths, so long as the configuration of the specimen and the illuminating system remained unaltered. The spectrophotometric setting would thus bear a constant ratio to the true values. Due to the presence of optical dispersion, this would hold only approximately for neighbouring wave-lengths. For example, the absorption $E_{0.44}$ of a specimen at 0.44μ could be measured by the orthodox method; that at 0.42μ would be too difficult to measure due to the low intensity of the source. It was possible, however, by using a much more powerful source, to obtain apparent values by the Pointolite method at the two wave-lengths, say $E'_{0.44}$ and $E'_{0.42}$, respectively. Then the true value $E_{0.42}$ at wave-length 0.42μ would not be far from that obtained by the calculation: $E_{0.42} = \frac{E'_{0.42}}{E'_{0.44}} \cdot E_{0.44}$. By this means the regions studied could be slightly extended. In the results presented, some of those at the extreme ends of the spectrum were obtained by the "overlapping method," and have been appropriately marked in the text. Usually at least two measurements at different wave-lengths were carried out to test the agreement with the normal method. It should, however, be noticed that this procedure may not be completely free from systematic errors. On the other hand, in the regions of low visibility the photometric accuracy depends largely on the brightness of the field in view. Furthermore, a field of greater brightness allows the use of filters to cut out stray light. With due care, the values so determined will be more reliable than those obtained by the usual method. But as it may be difficult to secure a uniform field by this method, care should be taken to obtain reasonable uniformity by determining the proper configuration of the specimen under test. Sometimes the insertion of a ground-glass screen between the specimen and the source may help in this respect, but the intensity of illumination would then be correspondingly lowered.

III. THE EXPERIMENTAL RESULTS.

1. The General Spectral Absorptions ‡ at Various Concentrations.

The results for the main series of glasses are set out in Table II. The specimens used were in the form of rods which had not been subjected to any heat-treatment after being drawn. In addition to the extinction coefficients at various wave-lengths, the specific extinction coefficient was also calculated, *i. e.*, the extinction coefficient per unit concentration of iron oxide. Further reference to this item will be made in a later section. In Table II the total amount of iron and the percentage dissociation, as well as the lengths or thicknesses of the specimens used, are also given. The data have also been plotted to give spectral absorption curves grouped according to the concentrations, in three separate graphs (figs. 1A-1C). The results obtained by the "overlapping method" at the two ends of the spectrum have been denoted by a "*" in the Table, while those marked with "†" are average values from both methods. In the graphs they have been distinguished by black dots and semi-black dots, respectively. In figs. 2A, 2B, and 2C, the extinction coefficients have been plotted against the iron oxide concentration at three particular wave-lengths, namely, $\lambda=0.45\mu$, 0.55μ , and 0.70μ , which correspond, approximately, to the blue, the green, and the red regions, respectively.

‡ Since at the time of writing, the task, undertaken by the Colour Group of the Physical Society, of preparing an agreed list of definitions relating to colour, reflexion, transmission, and absorption has not been completed, it seems desirable to define the terms employed in this paper. The terms "extinction coefficient" and "specific extinction coefficient" as employed by us are defined at the head of Table II. There remain two other terms to which reference should be made, namely, the "transmission" and the "transmittance."

"Transmission" is defined as the ratio of the intensity of the light emerging from the second glass surface to the intensity of that incident on the first surface.

"Transmittance" is defined as the ratio of the intensity of the light incident on the second surface to the intensity of that leaving the first surface.

These definitions are expressed mathematically as follows :

$$I=I_0 \left\{ 1 - \left(\frac{n-1}{n+1} \right)^2 \right\} 10^{-\alpha t} = I_0 \left\{ 1 - \left(\frac{n-1}{n+1} \right)^2 \right\} 10^{-\epsilon c t},$$

where I and I_0 are the intensities of light leaving the second surface and incident on the first surface, respectively, n is the refractive index, α the extinction coefficient, ϵ the specific extinction coefficient, c the percentage of colouring oxide present, and t the thickness of the specimen through which the light passes.

$$\text{Transmission} = I/I_0.$$

$$\text{Transmittance} = \frac{1}{I_0 \left\{ 1 - \left(\frac{n-1}{n+1} \right)^2 \right\}^2} = 10^{-\alpha}.$$

TABLE II.—The Extinction Coefficients and the Specific Extinction Coefficients of Soda-Lime-Silica Glasses containing varying amounts of Iron Oxide.

In this table, a =extinction coefficient in density per cm., ϵ_d =specific extinction coefficient in density per cm. per % iron oxide expressed as Fe_2O_3 . Data marked with “*” are results obtained by “overlapping method” (see text, p. 1). Data marked with “†” are averaged results obtained by both the normal and the “overlapping” methods.

Glass No. Total iron as Fe_2O_3 , FeO	Hypothetical blank.									
	A.1.	A.2.	A.3.	A.4.	A.5.	A.7.	A.1.	A.2.	A.3.	A.4.
Total iron (as Fe_2O_3) % Length of specimen (cm.)	0.002 %	0.022 %	0.032 %	0.042 %	0.053 %	0.077 %	0.002 %	0.022 %	0.032 %	0.042 %
λ (μ).	a	ϵ_d	a	ϵ_d	a	ϵ_d	a	ϵ_d	a	ϵ_d
0.41	—	—	—	—	—	—	—	—	—	—
0.42	—	—	—	—	—	—	—	—	—	—
0.43	0.0011	—	0.0056	—	0.0109	—	0.0109	—	0.0153*	—
0.44	—	—	0.0058	—	0.0108	—	0.0108	—	0.0174†	—
0.45	0.0019	—	0.0060	—	0.0105	—	0.0105	—	0.0171†	0.211
0.46	—	—	0.0045	—	0.0082	—	0.0082	—	0.0168†	0.208
0.47	0.0006	—	0.0037	—	0.0051	—	0.0051	—	0.0130	0.159
0.48	—	—	0.0037	—	0.0051	—	0.0051	—	0.0103	—
0.49	—	—	0.0037	—	0.0051	—	0.0051	—	0.0088	0.108
0.50	0.0004	—	0.0026	—	0.0034	—	0.0034	—	0.0088	—
0.51	—	—	0.0024	—	0.0029	—	0.0029	—	0.0073	—
0.52	—	—	0.0024	—	0.0029	—	0.0029	—	0.0068	0.086
0.53	—	—	0.0021	—	0.0029	—	0.0029	—	—	—
0.54	—	—	0.0021	—	0.0029	—	0.0029	—	—	—
0.55	0.0000	—	0.0023	—	0.0031	—	0.0031	—	—	—
0.56	—	—	0.0028	—	0.0037	—	0.0037	—	0.0052	0.107
0.57	—	—	0.0028	—	0.0037	—	0.0037	—	—	—
0.58	—	—	0.0034	—	0.0040	—	0.0040	—	—	—
0.59	—	—	0.0034	—	0.0040	—	0.0040	—	—	—
0.60	0.0000	—	0.0040	—	0.0053	—	0.0053	—	0.0089	—
0.61	—	—	0.0040	—	0.0053	—	0.0053	—	0.0122	0.180
0.62	—	—	0.0049	—	0.0067	—	0.0067	—	0.0154	—
0.63	—	—	0.0049	—	0.0067	—	0.0067	—	0.0136	—
0.64	—	—	0.0058	—	0.0081	—	0.0081	—	0.0189	—
0.65	0.0000	—	0.0058	—	0.0081	—	0.0081	—	0.0227	0.296
0.66	—	—	0.0067	—	0.0095	—	0.0095	—	—	—
0.67	—	—	0.0067	—	0.0095	—	0.0095	—	0.0192	—
0.68	—	—	0.0083	—	0.0114	—	0.0114	—	—	—
0.69	—	—	0.0083	—	0.0114	—	0.0114	—	0.0225	—
0.70	0.0000	—	0.0097	—	0.0126	—	0.0126	—	0.0320	0.452
0.71	—	—	0.0097	—	0.0143	—	0.0143	—	0.0455	—
0.72	—	—	0.0107	—	0.0155	—	0.0155	—	0.0374	—
0.73	—	—	0.0107	—	0.0155	—	0.0155	—	0.0273	0.531
0.74	—	—	0.0129	—	0.0177	—	0.0177	—	0.0518	—
0.75	—	—	0.0129	—	0.0177	—	0.0177	—	0.0306	—
0.76	—	—	0.0129	—	0.0177	—	0.0177	—	0.0310*	—
0.77	—	—	0.0129	—	0.0177	—	0.0177	—	—	—

TABLE II. (continued).

[illegible]

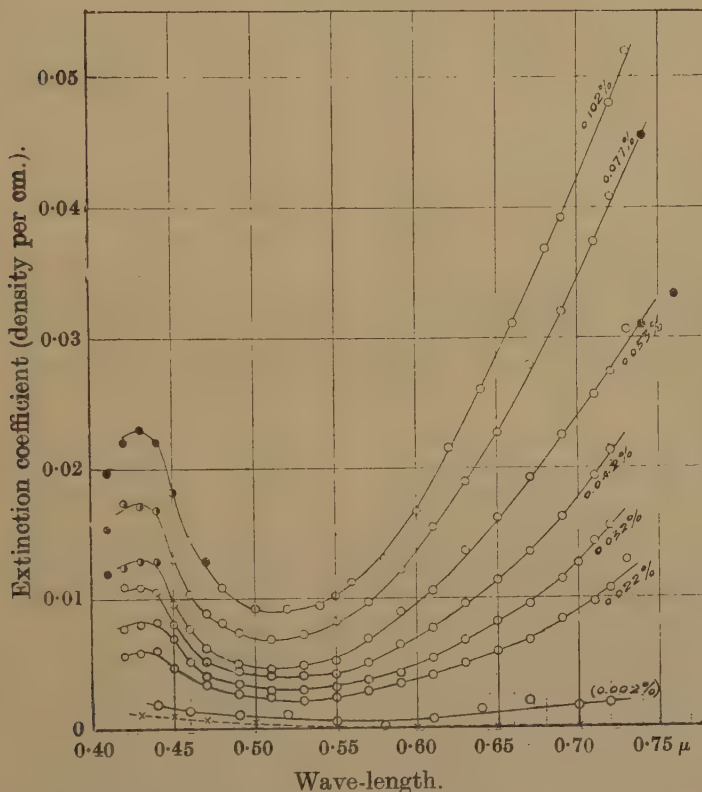
TABLE II. (continued).

Glass No. Total iron as Fe_2O_3 FeO	A.200.		A.300.		A.500.		A.650.		A.750.	
	1.90 %	ϵ_d^*	2.81 %	ϵ_d^*	4.67 %	ϵ_d^*	6.1 %	ϵ_d^*	6.7 %	ϵ_d^*
Total iron (as Fe_2O_3) %	11.4 %		10.3 %		9.2 %		8.5 %		8.9 %	
Length of specimen (cm.)	0.194		0.196		0.196		0.193		0.195	
λ (μ).	α .	ϵ_d^*	α .	ϵ_d^*	α .	ϵ_d^*	α .	ϵ_d^*	α .	ϵ_d^*
0.41	0.769*	—	1.56*	—	4.50*	—	8.00*	—	9.65*	—
0.42	0.784*	—	1.63*	—	4.08*	—	7.55*	—	8.91†	1.33
0.43	0.763*	0.401	1.62*	0.575	3.88*	—	6.92*	1.135	8.61†	1.14
0.44	0.795†	—	1.59	—	3.73*	—	6.68†	—	7.68†	—
0.45	0.588†	0.305	1.36†	0.484	3.42*	0.732	5.89†	0.965	6.42	—
0.46	—	—	—	—	2.72	—	—	—	5.60	—
0.47	0.392	0.206	0.919	—	2.41	—	4.28†	—	—	—
0.48	—	—	—	—	—	—	—	—	—	—
0.49	—	—	0.730	—	1.97	—	3.56	—	4.53	—
0.50	0.314	0.165	—	0.237	—	—	—	0.535	—	0.629
0.51	—	—	0.618	—	1.65	—	2.97	—	3.88	—
0.52	—	—	—	—	—	—	—	—	—	—
0.53	0.284	—	0.536	—	1.45	—	2.60	—	3.36	—
0.54	—	—	—	—	—	—	—	—	—	—
0.55	—	0.139	0.515	0.183	1.33	—	2.38	0.390	3.09	0.461
0.56	0.258	—	0.495	—	—	—	—	—	—	—
0.57	—	—	—	—	1.30	—	2.30	—	2.97	—
0.58	—	—	—	—	—	—	—	—	—	—
0.59	0.263	—	0.515	—	1.32	—	2.20	—	2.84	—
0.60	—	0.144	—	0.190	—	—	—	—	—	—
0.61	—	—	0.551	—	1.34	—	2.22	0.362	2.94	0.432
0.62	0.335	—	—	—	—	—	—	—	—	—
0.63	—	—	0.617	—	1.45	—	2.34	—	—	—
0.64	—	—	—	—	—	—	—	—	—	—
0.65	0.417	0.219	0.690	0.246	1.56	—	2.44	0.400	3.12	0.466
0.66	—	—	—	—	—	—	—	—	—	—
0.67	—	—	0.750	—	1.60	—	2.56	—	—	—
0.68	0.624	—	0.806*	—	1.72*	—	2.64*	—	—	—
0.69	—	—	0.856	—	1.84	—	2.74	—	—	—
0.70	0.737	0.387	0.944	0.336	1.93*	0.413	2.84	0.465	3.53†	0.526
0.71	—	—	1.01	—	2.07†	—	—	—	3.80†	—
0.72	0.863†	0.455	1.12*	0.399	2.15*	0.460	3.20	0.525	4.11†	0.614
0.73	—	—	—	—	—	—	—	—	4.66*	—
0.74	0.940*	—	1.25*	—	—	—	—	—	—	—
0.75	—	—	—	—	2.36*	—	—	—	—	—

2. The Correction for Platinum in Solution.

Although precautions had been taken to reduce the solution of platinum from the glass-melting crucible to the minimum by sintering the batch before melting, the platinum dissolved during the 22 hours of melting at 1400° was still quite sufficient to cause a measurable absorption. This fact becomes apparent from fig. 2A; for extrapolation of the absorption curves to zero concentration of iron oxide, particularly the one at $\lambda = 0.45 \mu$, does not give zero absorption. The same was true

Fig. 1A.

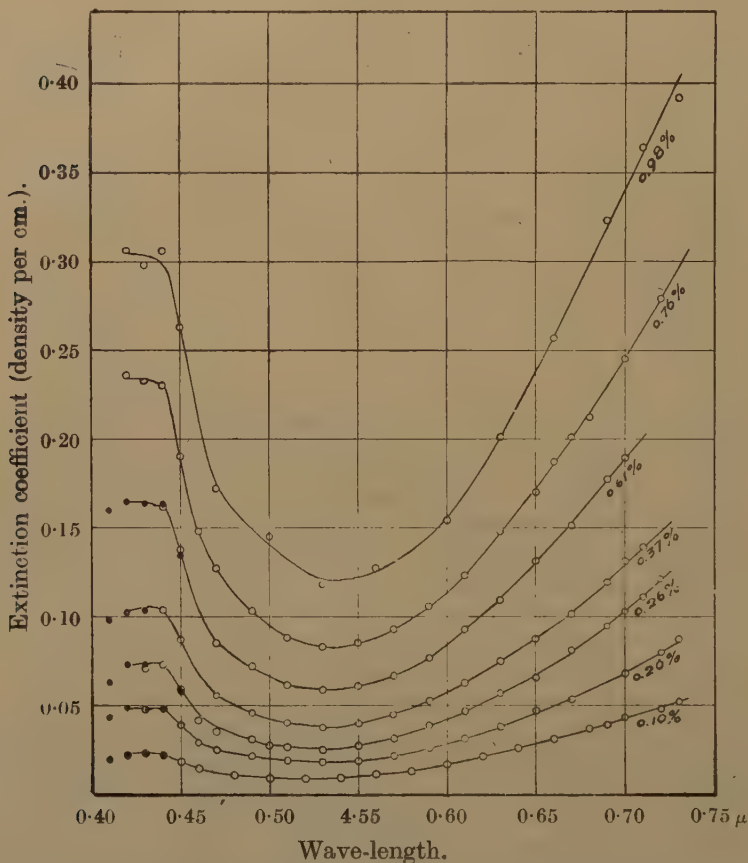


Extinction coefficients at different wave-lengths of soda-lime-silica glasses containing 0.002–0.102% of iron oxide expressed as Fe_2O_3 .

when the absorptions for all wave-lengths in the blue-violet region were plotted in the same way. These residual spectral absorptions are plotted in fig. 1A near the abscissa as a dotted curve. Strictly speaking, we are not quite certain that these residual absorptions were due to platinum absorption alone. But from other evidence obtained by the authors in their recent melts, it is believed that platinum was mainly responsible for the additional absorption. Furthermore, there was evidence that the amount of platinum in solution depended also on the iron concentra-

tion. It seemed, however, reasonable to assume that at high iron oxide concentrations the effect of platinum on the absorption would be relatively small. For low concentrations, say below 0.1 per cent. of Fe_2O_3 , the amount in solution may be regarded as remaining practically constant. As this residual spectral absorption was fairly small in comparison with the absorptions studied, it was taken as a blank or standard. In the subsequent evaluation of the specific extinction coefficients, allowance was made for these blank values in every case.

Fig. 1B.



Extinction coefficients at different wave-lengths of soda-lime-silica glasses containing 0.10–0.98% of iron oxide expressed as Fe_2O_3 .

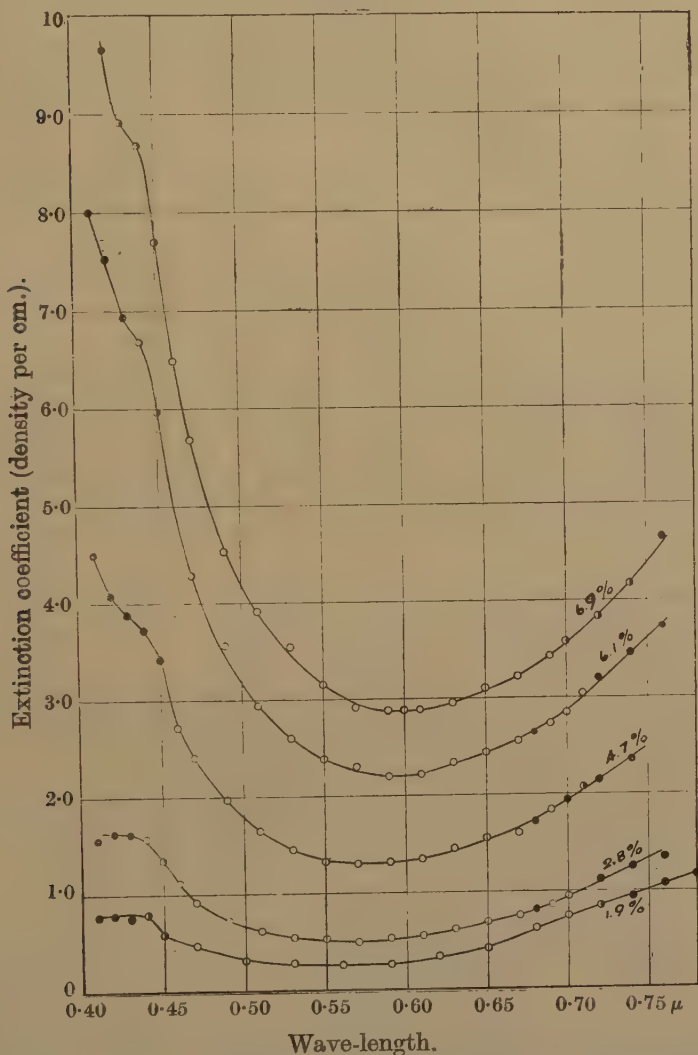
3. The General Trend of the Absorption in Relation to Concentration.

Figs. 1A–1C, and 2A and 2B reveal the following facts:

(a) All the absorption curves (figs. 1A–1C) show a minimum in the green region. This is a well-known fact. It is sufficient to note that the minimum shifts gradually from $\lambda=0.52\mu$ to $\lambda=0.60\mu$ as the concentration of iron oxide increases from zero to 6.7 per cent.,

(b) At low concentrations there is a "plateau" of absorption in the blue-violet region $\lambda=0.43\ \mu$. This gradually loses its prominent shape for concentrations above 2.8 per cent. and at 6.7 per cent. takes the form of a slight kink only.

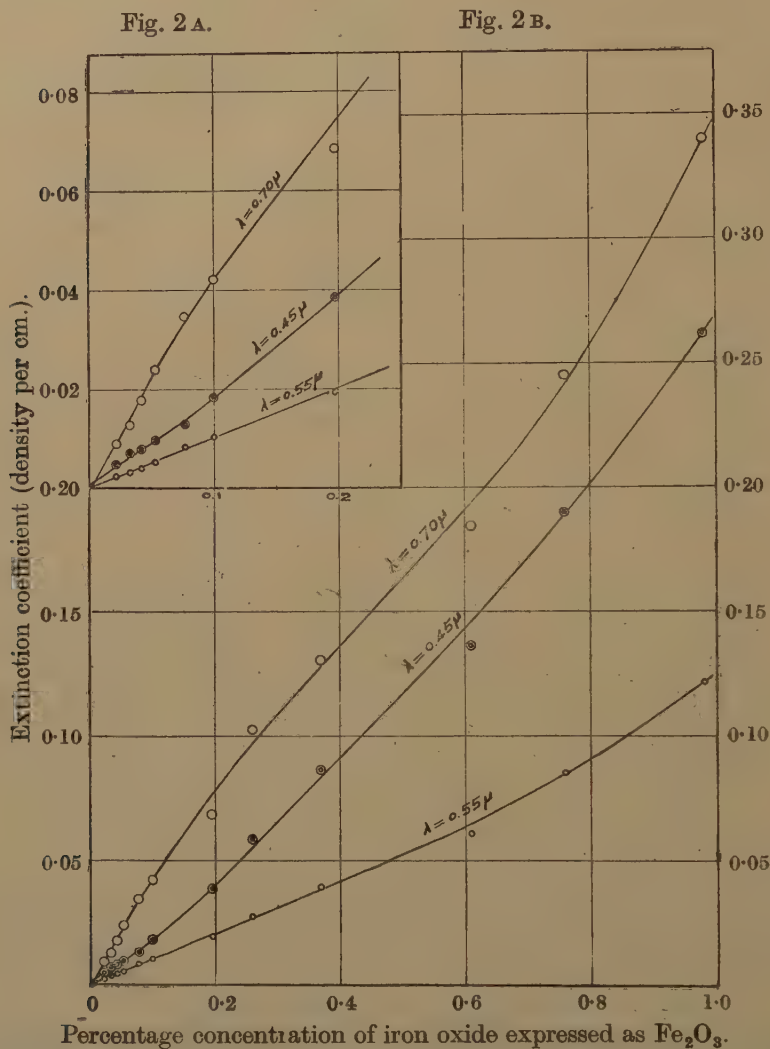
Fig. 1 c.



Extinction coefficients at different wave-lengths of soda-lime-silica glasses containing 1.9-6.9% of iron oxide expressed as Fe_2O_3 .

Despite the fact that the accuracy attainable in this region was low, the existence of the above feature can be regarded as definite in view of the particular precautions taken. First, the photometer scale was calibrated specially at that wave-length in order to ensure no abnormal

performance. Secondly, as has been mentioned above, the instrument was free from observable systematic errors with regard to stray light. This was further proved by the fact that the insertion of additional filters in the beam made no difference to the results observed.



Relation between the extinction coefficient at some particular wave-length and iron oxide concentration.

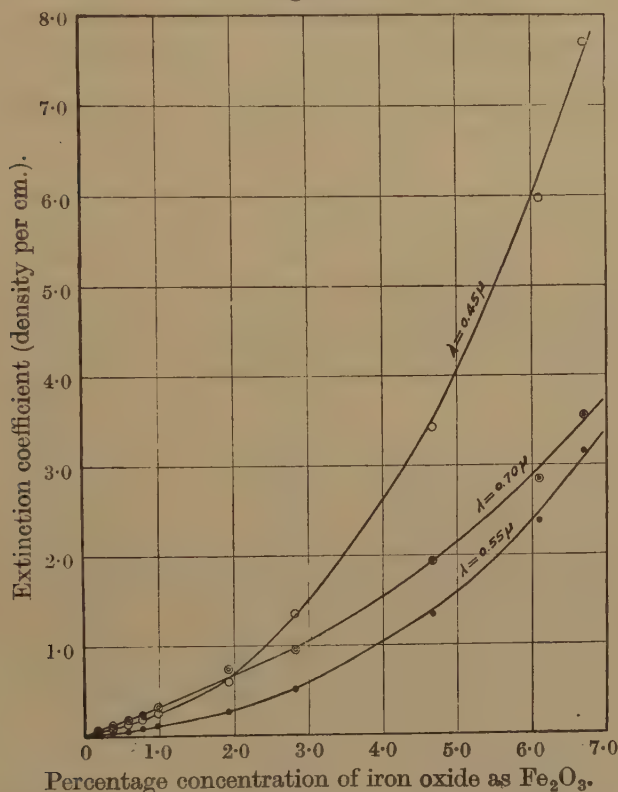
Actually, the plateau occurs in many types of glasses containing iron impurities. In the literature, mainly on optical glasses, such as borosilicate crown, crown, light flint, etc., one finds that absorption in this region shows either a plateau or a maximum*. The results of an

* See *Glastechnische Tabellen*, published by J. Springer, Berlin, 1932.

investigation by C. Andresen-Kraft* on soda-silica glasses with comparatively high iron oxide concentrations also showed this effect, but it was not specifically mentioned by the author. The plateau, however, is revealed if the results quoted by Andresen-Kraft for the spectral absorption curves in the U.V. and the visual region are plotted on the same graph.

(c) The absorption at all wave-lengths increased proportionately with increasing iron oxide content up to 0.1 per cent. (figs. 2 A and 2 B). Then a bend appeared in the curves corresponding both to the red and the

Fig. 2c.



Relation between the extinction coefficient at some particular wave-length and iron oxide concentration.

violet regions. This, as will be seen later, is directly connected with the sharp change of percentage dissociation in this concentration range. The absorption curves then became linear again, up to 0.6 per cent., where a tendency towards a general increase in absorption could be seen. This increase continued with the iron oxide concentration. The

* *Glastech. Ber.* xi. p. 577 (1931).

greatest change in slope of the absorption curve seemed to appear between 3 and 4 per cent. Fe_2O_3 .

(d) In conjunction with the features described in (c), increasing the concentration of iron oxide when the total Fe_2O_3 was still comparatively low produced an absorption in the red region greater than in the blue-violet region, the reverse being true at higher concentrations.

Further discussion of these points will be taken up later.

4. The Correlation between Percentage Dissociation and Specific Extinction Coefficient.

In figs. 3A and 3B the specific extinction coefficients are plotted at chosen wave-lengths covering the whole range of the visual spectrum.

Fig. 3A

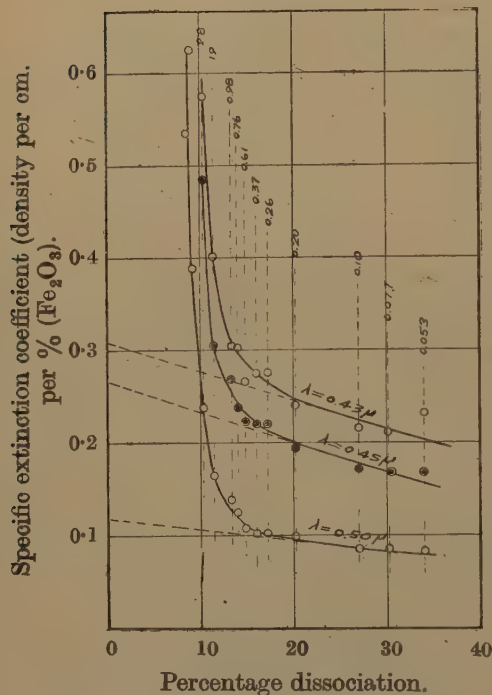
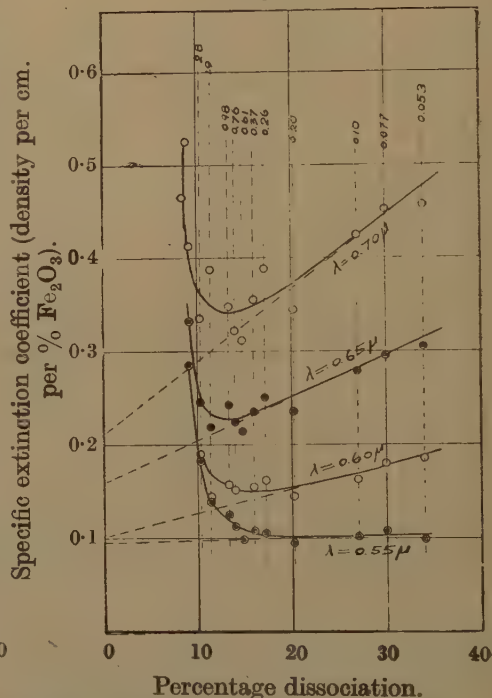


Fig. 3B.



(Figures quoted in the graphs are percentage concentrations of total iron as Fe_2O_3 .)

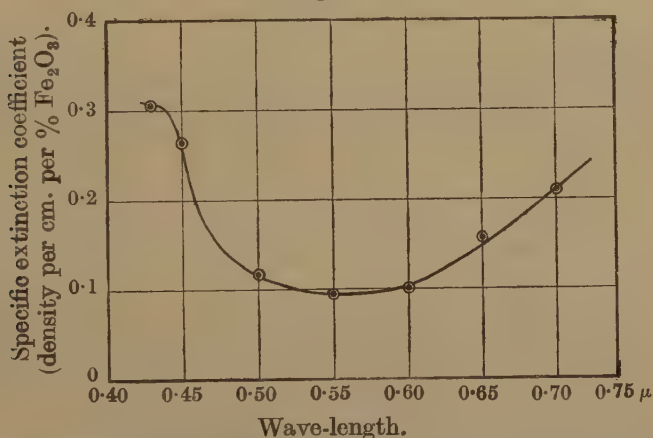
Correlation between absorption of iron oxide and ferrous-ferric dissociation.

Some of the specific extinction coefficients are interpolated values from the spectral absorption curves, as can be noticed from Table II. In examining the graphs one fact that should be remembered is that all the points are obtained from different concentrations. Each point is therefore appropriately marked by its iron oxide content. Owing to the limited accuracy attainable at extremely low concentrations, the

results for the glasses containing less than 0.05 per cent. were not included.

From the graphs for glasses of iron oxide content below 0.6 per cent. expressed as Fe_2O_3 , there appears to be a linear relation between percentage dissociation and specific extinction coefficient. It seems reasonable to assume that if there were no variation in dissociation, the specific extinction coefficient would have kept constant up to the concentration of 0.6 per cent. To explain the linear relation which covers a range in dissociation of about 20 per cent., it may be suggested that a simple exchange relation exists between the ferric and the ferrous iron, although it is impossible to indicate whether or not there is a joint effect of the two oxides of iron, FeO and Fe_2O_3 . One thing is certain, and is, of course, well known—namely, that with a higher percentage of dissociation the absorption in the red region increases greatly, while the blue-violet absorption decreases only moderately. The greater

Fig. 4.



Result of extrapolation from figs. 3A and 3B. Hypothetical spectral absorption curve of soda-lime-silica glass (75 % SiO_2 , 15 Na_2O , 10 CaO) containing low amounts of iron oxide in 100 % ferric state.

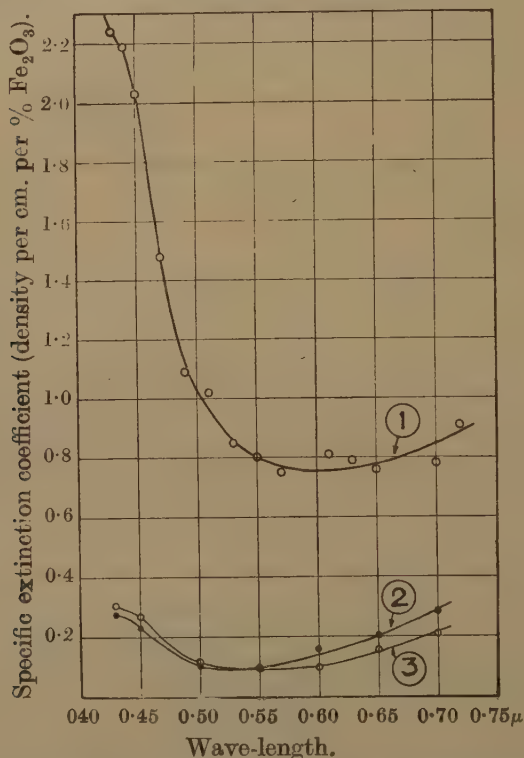
fluctuations of results in the red region may also indicate that the ferrous iron may be very sensitive to heat-treatment.

Owing to the fact that the lowest degree of dissociation for the glasses under consideration is only 15 per cent., a linear extrapolation towards the ferric side should yield useful information. Fig. 4 shows the result of such extrapolation which represents hypothetically the spectral absorptions of soda-lime-silica glass containing low concentrations of iron oxide in the state of 100 per cent. Fe_2O_3 . We may here stress the fact that in soda-lime-silica glass of the composition cited, the effect on the spectral absorption of low concentrations of iron oxide in the ferric state differs decidedly from that at high concentrations. Conclusions about the effect of high iron oxide concentrations have hitherto

been drawn* from tests on glasses melted from batches containing strong oxidizing agents, such glasses showing strong absorption in the blue-violet region and a gradual decrease in absorption towards the red end. The observed colours were either brown or amber or yellowish-brown, according to the concentration and thickness of the specimen. At low concentrations the red region was marked by considerable absorption, and consequently the colour shows up as yellowish-green.

Fig. 5.

A comparison of spectral absorptions at high and low concentrations of iron in soda-lime-silica glass.



Curve 1. Rate of change of absorptions with iron concentration near 6.7 % iron expressed as Fe_2O_3 . 8.9 % ferrous-ferric dissociation.

Curve 2. Absorptions for low concentrations of iron at 8.9 % ferrous-ferric dissociation curve derived from fig. 3 A and 3 B.

Curve 3. Absorptions for low concentrations of iron at 100 % ferric state. (Same curve as fig. 4.)

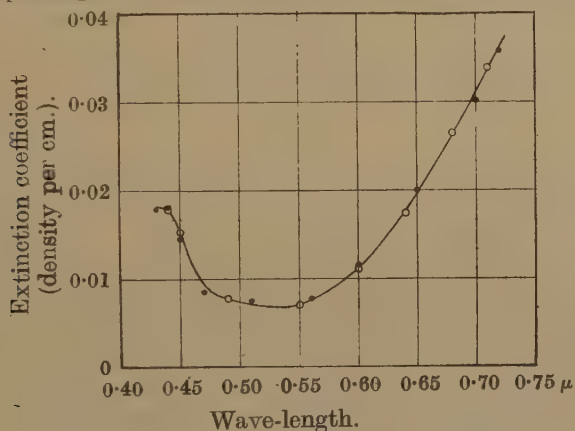
Furthermore, the specific extinction coefficients are in general much lower than at high concentrations.

* See, for example, W. E. S. Turner and W. Weyl, *J. Soc. Glass Tech.*, Trans. xix. p. 209 (1935); W. Weyl, *Sprechsaal*, lxxi. pp. 91, 104, and 117 (1938).

Returning to figs. 3 A and 3 B again, as the concentration of iron oxide rose above 0.6 per cent., the specific extinction coefficient increased very rapidly, particularly in the blue-violet region. Here again is evidence that the absorption characteristics at high concentrations are governed by some reaction quite different from that at low iron oxide content, even though the ferrous iron may have influenced the results. This difference is clearly seen from fig. 5, in which the curve for low ferric iron concentration is contrasted with one for the glass containing 6.7 per cent. A spectral absorption curve corresponding to low concentration, but at the same percentage dissociation as at 6.7 per cent. total iron oxide, is also shown to facilitate comparison. It may be noted that the "high concentration effect" began to be noticeable at concentrations as low as 0.26 per cent. This being so, it is clearly important that the study of the processes of decolorizing glasses should be based on specimens containing concentrations of iron oxide falling within, or not greatly exceeding, the limiting amount of iron oxide which can be decolorized.

Fig. 6.

Effect of prolonged melting on glass containing 0.075 % iron oxide.



In the graph : ○ melted 24 hrs. at 1400°

● " 100 " "

5. The Effect of Prolonged Heating and Reheating.

In the previous paper by Densem and Turner, the conclusion was drawn that under the conditions of the experimental work—namely, melting electrically at 1400° in contact with the atmosphere—the degree of dissociation in a glass containing 0.075 per cent. total iron oxide was the same after 100 hours' heating as after 24 hours. Neither was there a change observable on reheating rods of this glass for 24–48 hours at 500–600°. Samples of this glass after such treatments were examined spectrophotometrically. The results are shown in fig. 6, which clearly indicates that such prolonged heating made no difference to the spectral absorption. The results of measurements on the glass containing 0.26

per cent. in the main series showed (fig. 7A) that no significant difference could be found between a specimen which had not been annealed and

Fig. 7A.

Effect of heat-treatment on glass containing 0.26 % iron oxide ("low" concentration).

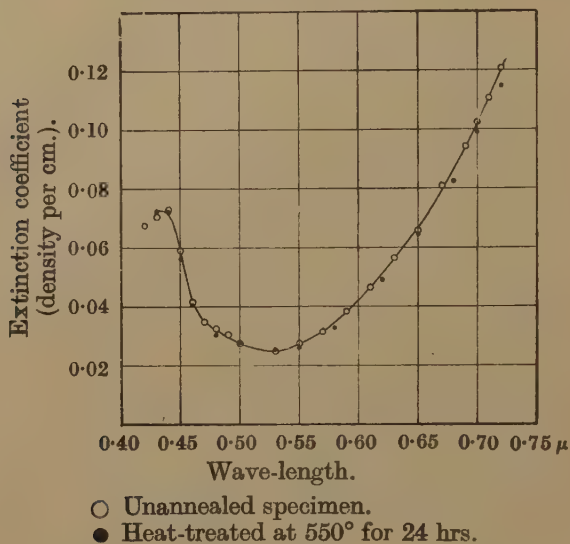
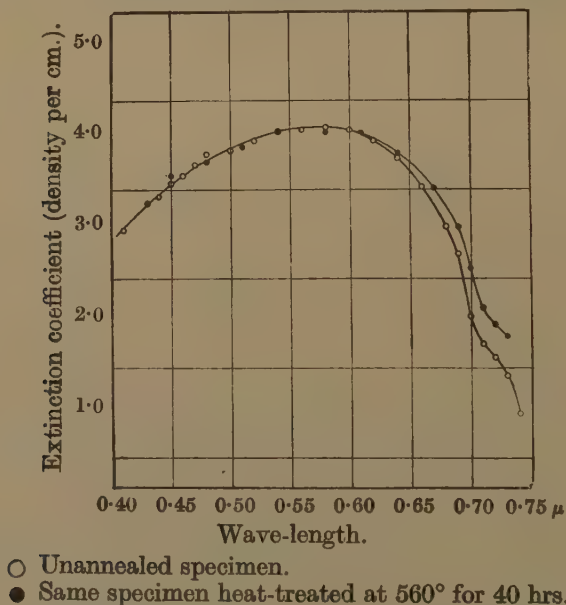


Fig. 7B.

Effect of heat-treatment on glass containing 4.7 % iron oxide.



another reheated at 550° for 24 hours. On the other hand, the spectral absorption of a specimen of the glass containing 4.7 per cent. iron oxide,

before and after heat-treatment at 560° for 40 hours, showed (fig. 7 B) a difference in the blue-violet region, the heat-treated specimen having a lower absorption. These results again support the view that the action when high concentrations of iron oxide are present differs from that at low concentrations.

IV. GENERAL DISCUSSION AND CONCLUSIONS.

The main results of this investigation can be summarized as follows :

(1) At low concentrations of iron oxide, say, below 0.3 per cent. expressed as Fe_2O_3 , there is a marked rise in the dissociation of Fe_2O_3 , and the increased FeO concentration is responsible for considerable absorption in the red region, enhanced by a moderate reduction of the ferric absorption. In other words, below the concentration equivalent to 0.2 per cent. Fe_2O_3 , the ferrous part exercises the greater colouring power, which may be regarded as blue.

(2) Extrapolation of the results indicates that at low concentrations iron oxide, when completely in the ferric state, gives a yellow or greenish-yellow colour to the soda-lime-silica glasses of the compositions quoted, due to the fact that considerable absorption occurs in both the blue and the red regions.

(3) At concentrations above 0.6 per cent. the absorption takes a further form. Assuming this second form of absorption as due to ferric iron, it is, in general, much greater than that at lower concentrations, the most marked absorption being in the blue-violet region of the visible spectrum, and very probably extending right into the ultra-violet region.

(4) The highly absorptive form of ferric oxide is sensitive to heat-treatment. At low concentrations, where the low absorptive component prevails, no observable effect has been found after heat-treatment near the softening point.

So far, much more has been said about the behaviour of ferric iron in the glass, because the iron oxide dissociation percentages were all much nearer the ferric than the ferrous side and their significance is more easily understood. Although we can say that the ferrous iron contributes a blue colour, we are unable to draw a definite conclusion, based on the above observations alone, whether this blue colour arises from a simple action or a complex one. The linear relation between absorption and ferric-ferrous equilibrium should no doubt be considered as only an approximation. Some preliminary experiments were made on glasses containing low concentrations of iron oxide melted in a gas-fired furnace under reducing conditions. The percentage dissociation was on the high side, and in all cases the absorption in the red region was higher than that expected by the linear extrapolation towards high dissociation. Probably the absorption due to ferrous iron is very susceptible to heat-treatment. Alternatively, the function of absorption may be fairly

complex. Unfortunately, when the present work was in progress, circumstances did not allow us to make melts under careful control in an electric furnace, and reproducible melts would have to be obtained in order to clarify this point. And when this doubt has been settled it has to be borne in mind that all the results apply to one particular soda-lime-silica glass, and the study would need to be made of the other glasses used by Densem and Turner—namely, the lithia-silica, soda-silica, and potash-silica glasses—whilst the lead oxide-silica glasses should also be included.

From the theoretical point of view, we are familiar with the fact of oxidation and reduction; and the existence of different absorption characteristics of colouring agents in glass, even when the state of valency remains unchanged, is well known. It is influenced by the atomic fields exerted by the surrounding atoms on the colouring ions. A striking example of this phenomenon is the behaviour of nickel oxide in glass. W. Weyl and E. Thümen*, in comparing vitreous solutions with aqueous solutions, concluded that nickel is present in glass either in the ionic state or in close association with other atoms, such that it behaves like an undissociated "molecule." The effect of heat-treatment, whereby the absorption characteristics of one phase were increased at the expense of the other, was regarded as showing that more association was taking place during the period of slow cooling. It may be that ferric iron in glass behaves in a more or less similar manner. We have noted that at low concentrations the percentage dissociation of iron oxide tends to approach a constant value of about 31 per cent. (Table I.). Starting from 0.1 per cent. of iron oxide, this oxidation-reduction percentage begins to fall gradually, until when 4 per cent. iron oxide is present the dissociation value tends to become constant again at about 9.5 per cent. It may be that the percentage dissociation at very low concentrations of iron oxide corresponds to equilibrium with one form of associated ferric iron, whilst the 9.5 per cent. dissociation corresponds to another state of association of ferric iron prevailing only at the comparatively high concentrations. In regard to the absorption curves, the linear relation holds good for low concentrations of iron oxide up to 0.6 per cent., although a trace of the high absorptive component can be perceived at as low as 0.2 per cent. iron oxide (figs. 3A and 3B). At high concentrations the greatest change in the shape of the absorption curve occurred also at approximately 3-4 per cent. Bearing in mind the fact that the high concentration absorption component of ferric iron is sensitive to heat-treatment, whilst the low concentration component is not as sensitive, the authors feel justified in regarding the low-absorption component as exhibiting a more stable molecular association.

In support of the above view there may be cited the result of the investigation carried out by E. Preston and W. E. S. Turner on the effect of small amounts of various colouring oxides on the colour of a

* *Sprechsaal*, lxv. p. 658 (1932).

soda-lime-silica glass*. A glass of parent composition 73.5 SiO₂, 10 CaO and 16.5 per cent. Na₂O, containing in addition 0.05 per cent. iron oxide added as Fe₂O₃, was melted in an electric furnace at 1400° and specimens were taken at intervals. $\frac{1}{2}$, 1, and 3 hours after the glass had become seed-free. The specimen taken after $\frac{1}{2}$ hour heating showed a decidedly stronger brown tint than the one taken after 3 hours, which had a yellowish-green colour. This observation could be interpreted in the light of the results of our present work by assuming that when the glass was first melted, the ferric iron was present in predominance in the "high concentration" form, leading to an amber or brown tint. As the heating was prolonged, not only did some dissociation from ferric to ferrous iron take place, but also the undissociated ferric iron became gradually more firmly associated molecularly, leading to the lighter-tinted yellowish-green colour. That it should take 3 hours to change from one colour to the other would be evidence that the association takes considerable time to complete.

It has already been noted that for glasses containing small amounts of iron, after being subjected to solarization or exposure under ultra-violet light, some of the ferrous iron becomes oxidized to the ferric state†. It can be seen from the results of A. E. Badger and A. C. Ottoson‡, in their studies on the effect of light on the colour of glass, as also from a study by W. M. Hampton and W. N. Wheat on optical glasses§, in which the spectral absorption of a boro-silicate crown glass was given both before and after exposure under ultra-violet light, that the change of absorption characteristics did not follow a simple exchange relation between ferrous and ferric iron. Although the discrepancies may be mainly due to the presence of arsenic, which alone gives a solarization effect, one must also bear in mind the possible formation of ferric iron in the less associated state. One can suppose that the ferric iron formed by the photo-electric effect would not fit in too well with the surrounding atoms unless some further heat-treatment was given.

In dense flint glasses containing a high percentage of lead oxide it has been shown|| that the colouring power of iron is much stronger than that in an ordinary soda-lime-silica glass. If we examine the absorption characteristics of the lead-containing glasses, where a remarkably high absorption occurs in the violet end of the visible spectrum, one may probably regard the ferric iron in this case as less associated. Presumably the surrounding atoms do not favour the saturated co-ordination of ferric ions.

* J. Soc. Glass Tech., Trans. xv. p. 5 (1941).

† D. Starkie and W. E. S. Turner, J. Soc. Glass Tech., Trans. xii. p. 306 (1928).

‡ J. Amer. Ceram. Soc. xxv. p. 104 (1942).

§ J. Soc. Glass Tech., Trans. xv. p. 306 (1931).

|| Phys. Soc. Reports, v. p. 185 (1938).

Finally, the authors wish to point out that the values of the thermal expansion and the viscosity near the softening point previously measured * in the Department for a similar series of glasses all show a marked bend at an iron-oxide concentration of about 3-4 per cent. Although, owing to the complications involved in this subject, we cannot yet be certain that these data can be correlated with the present results, the phenomenon is worthy of note.

The absorption curves determined by Andresen-Kraft for soda-silica glasses containing iron mainly in the form of Fe_2O_3 seem also to indicate the presence of two separate absorptive components, one corresponding essentially to the low concentration form of iron and the other component the high concentration form.

ACKNOWLEDGMENT.

In conclusion, the authors wish to express their gratitude to Miss V. Dimbleby and her colleagues of the chemical laboratory who carried out the chemical analyses. One of us (T. H. Wang) is also indebted to the Glass Delegacy for the Research Fellowship which enabled him to carry on the present work.

Department of Glass Technology,
The University, Sheffield.

LXXXVII. *Some Boundary Problems of Two-Dimensional Elasticity.*

By A. C. STEVENSON †.

[Received April 27, 1943.]

THIS paper is concerned with the setting up of boundary equations in problems of two-dimensional elasticity in terms of two complex potentials and their solution in certain simple but interesting cases by elementary tentative methods. Some of the solutions are equivalent to the Cauchy integral solutions for the corresponding analytic functions which would follow from Muschelishvili's method, which only recently became known to the writer. A minor but non-essential difference of presentation is that whereas Muschelishvili's approach to the problem is by way of the Airy stress function, the present treatment by complex potentials finds no explicit need for the Airy stress function. This method of complex potentials has also been developed from the beginning to cope with problems involving body-force, and the present examples of the method include some of gravitational body-force and of the reversed mass-accelerations in steady rotation treated as body-forces.

* F. R. Harris, *Glass Review*, xiv. pp. 110 and 128 (1938).

† Communicated by the Author.

1. The Differential Equations for the Displacements.

Using the co-ordinates z, \bar{z}, Z instead of the cartesian co-ordinates x, y, Z , where $z=x+iy, \bar{z}=x-iy$ (and generally using a bar over a complex quantity to denote the conjugate complex quantity), together with the abbreviations Θ, Φ, Ψ for the stress combinations

$$\Theta = \widehat{xx} + \widehat{yy}, \quad \Phi = \widehat{xx} - \widehat{yy} + 2i\widehat{xy}, \quad \Psi = \widehat{xz} + i\widehat{yz}, \quad \dots \quad (1.1)$$

the body-stress equations are

$$\frac{\partial \Phi}{\partial z} = \frac{\partial}{\partial \bar{z}} (\Theta - 2\rho U) + \frac{\partial \Psi}{\partial Z} = 0, \quad \frac{\partial \Psi}{\partial z} + \frac{\partial \bar{\Psi}}{\partial \bar{z}} + \frac{\partial}{\partial Z} (\bar{z}z - \rho U) = 0, \quad (1.2)$$

where $U(z, \bar{z}, Z)$ is the body-force potential function and ρ is the density of the elastic material.

With the usual notation u, v, w for the cartesian displacements, we use the abbreviation $D = u + iv$, and the complex form of the stress-displacement relations corresponding to the usual cartesian stress-strain relations in terms of the displacements, are

$$(1-2\eta)\Theta = 2\mu \left\{ \frac{\partial D}{\partial z} + \frac{\partial \bar{D}}{\partial \bar{z}} + 2\eta \frac{\partial w}{\partial Z} \right\}, \quad \Phi = 4\mu \frac{\partial D}{\partial \bar{z}}, \quad \dots \quad (1.3)$$

$$(1-2\eta)\bar{z}z = 2\mu \left\{ \eta \left(\frac{\partial D}{\partial z} + \frac{\partial \bar{D}}{\partial \bar{z}} \right) + (1-\eta) \frac{\partial w}{\partial Z} \right\}, \quad \Psi = \frac{\partial D}{\partial Z} \mu + 2\mu \frac{\partial w}{\partial \bar{z}}, \quad (1.4)$$

where Poisson's ratio η is related to the Lamé constants λ, μ by the equation

$$\eta = \lambda / 2(\lambda + \mu). \quad \dots \quad (1.5)$$

Eliminating the stresses from (1.2) by means of (1.3) and (1.4), we have

$$\frac{\partial^2 w}{\partial \bar{z} \partial Z} + (3-4\eta) \frac{\partial^2 D}{\partial z \partial \bar{z}} + \frac{\partial^2 \bar{D}}{\partial \bar{z}^2} + \frac{1}{2}(1-2\eta) \left\{ \frac{\partial^2 D}{\partial Z^2} - 2\frac{\rho}{\mu} \frac{\partial U}{\partial \bar{z}} \right\} = 0, \quad (1.6)$$

$$\frac{\partial}{\partial Z} \left(\frac{\partial D}{\partial z} + \frac{\partial \bar{D}}{\partial \bar{z}} \right) + 2(1-\eta) \frac{\partial^2 w}{\partial Z^2} + (1-2\eta) \left\{ 4\frac{\partial^2 w}{\partial z \partial \bar{z}} - \frac{\rho}{\mu} \frac{\partial U}{\partial Z} \right\} = 0, \quad (1.7)$$

as the differential equations satisfied by the displacements D, w in the complex co-ordinate system z, \bar{z}, Z .

2. The Complex Beltrami Equations (with Body-Force).

With the device of the Kronecker delta ($\delta_{pq}=1$ for $p=q$, but $=0$ for $p \neq q$), the equations of Beltrami generalized to include body-forces derivable from a potential function

$$V(x, y, Z) = U(z, \bar{z}, Z) \quad \dots \quad (2.1)$$

can be written in the single form

$$\nabla^2 \{ (1+\eta) \widehat{pq} - \eta (\Theta + \bar{z}z) \delta_{pq} \} + \frac{\partial^2}{\partial p \partial q} \{ \Theta + \bar{z}z - 2\rho(1+\eta)V \} = 0, \quad (p, q = x, y, z). \quad (2.2)$$

These six equations are obtained from the cartesian stress-strain relations equivalent to (1.3) and (1.4) and the body-stress equations equivalent to (1.2), and it has been shown by Filon⁽¹⁾ that stresses satisfying (1.2) and (2.2) must be compatible and form a possible solution of the stresses in an elastic solid under the given body-force. See also Seth⁽²⁾.

Abbreviate by writing

$$(1+\eta)B=\Theta+\widehat{z\bar{z}}-2\rho(1+\eta)U, \quad \nu=\eta/(1-\eta), \quad \nabla^2=4\frac{\partial^2}{\partial z\partial\bar{z}}+\frac{\partial^2}{\partial Z^2}, \quad (2.3)$$

and it is readily found that the complex form of the Beltrami equations is given by

$$\nabla^2\{\Theta-2\rho\nu U\}+4\frac{\partial^2 B}{\partial z\partial\bar{z}}=0, \quad \nabla^2\{\widehat{z\bar{z}}-\rho\nu U\}+\frac{\partial^2 B}{\partial Z^2}=0, \quad (2.4)$$

$$\nabla^2\Phi+4\frac{\partial^2 B}{\partial z^2}=0, \quad \nabla^2\Psi+2\frac{\partial^2 B}{\partial\bar{z}\partial Z}=0. \quad (2.5)$$

The equations (2.5) being complex, are, with their conjugate complexes, equivalent to four, bringing the Beltrami equations up to six in number as in (2.2).

3. Plane Strain.

Plane strain is defined as the case

$$\frac{\partial D}{\partial Z}=0, \quad w=0, \quad \frac{\partial U}{\partial Z}=0, \quad \dots \quad (3.1)$$

which makes

$$(1-2\eta)\Theta=2\mu\left(\frac{\partial D}{\partial z}+\frac{\partial\bar{D}}{\partial\bar{z}}\right), \quad \widehat{z\bar{z}}=\eta\Theta, \quad \Phi=4\mu\frac{\partial D}{\partial\bar{z}}, \quad \Psi=0, \quad (3.2)$$

where from (1.6), D satisfies

$$\frac{\partial}{\partial\bar{z}}\left\{\kappa\frac{\partial D}{\partial z}+\frac{\partial\bar{D}}{\partial\bar{z}}-\frac{1}{2}\frac{\rho}{\mu}(\kappa-1)U\right\}=0, \quad \kappa=3-4\eta, \quad (3.3)$$

which gives, on integration with respect to \bar{z}

$$\kappa\frac{\partial D}{\partial z}+\frac{\partial\bar{D}}{\partial\bar{z}}-\frac{1}{2}\frac{\rho}{\mu}(\kappa-1)U=f(z)=C\Omega'(z),$$

introducing a real constant C and an arbitrary function of z in the differentiated form for later convenience.

This, and the conjugate complex equation, are linear simultaneous equations for $\partial D/\partial z$ and its conjugate, and writing

$$C=(\kappa^2-1)/8\mu, \quad U(z, \bar{z})=\frac{\partial W}{\partial z}(z, \bar{z}), \quad \gamma=4(\kappa-1)/(\kappa+1), \quad (3.4)$$

their solution for $\partial D/\partial z$ can be written

$$8\mu\frac{\partial D}{\partial z}=\kappa\Omega'(z)-\bar{\Omega}'(\bar{z})+\gamma\rho\frac{\partial W}{\partial z},$$

so that, integrating with respect to z and introducing a second arbitrary function, again in a differentiated form for later convenience, we have

$$8\mu D = \kappa \Omega(z) - z \bar{\Omega}'(\bar{z}) - \bar{\omega}'(\bar{z}) + \gamma \rho W, \quad (3.5)$$

giving a result obtained elsewhere in this form⁽³⁾ (1940) by a different route by the present writer, and equivalent to results obtained much earlier by Kolosof⁽⁴⁾ (1908).

From (3.2), we find for the stresses

$$2\Theta = \Omega'(z) + \bar{\Omega}'(\bar{z}) + (4 - \gamma)\rho \frac{\partial W}{\partial z}, \quad -2\Phi = z \bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z}) - \gamma \rho \frac{\partial W}{\partial \bar{z}} \quad (3.6)$$

in terms of the functions $\Omega(z)$, $\omega(z)$ (which we term the *complex potentials* of plane strain), and the elastic constant γ .

4. Generalized Plane Stress.

Filon's method of reducing certain types of problems relating to plate-like material, bounded by planes $Z = \pm c$, to two-dimensional problems analytically almost the same as that of plane strain, is to define a state of *generalized plane stress* as given by

$$\partial U / \partial Z = 0, \quad \bar{z} \bar{z} = 0 \text{ everywhere,} \quad \bar{x} \bar{z} = 0 \text{ over } Z = \pm c, \quad (4.1)$$

and attention is then focussed on the *mean stresses* $\bar{x}\bar{x}$, $\bar{y}\bar{y}$, $\bar{x}\bar{y}$ and *mean displacements* \bar{u} , \bar{v} defined, in complex combination, by

$$\Theta_0 = \bar{x}\bar{x} + \bar{y}\bar{y}, \quad \Phi_0 = \bar{x}\bar{x} - \bar{y}\bar{y} + 2i\bar{x}\bar{y}, \quad D_0 = \bar{u} + i\bar{v}, \quad (4.2)$$

$$\text{and} \quad \Theta_0 = \frac{1}{2c} \int_{-c}^c \Theta dZ, \quad \Phi_0 = \frac{1}{2c} \int_{-c}^c \Phi dZ, \quad D_0 = \frac{1}{2c} \int_{-c}^c D dZ. \quad (4.3)$$

Then, from (1.2), we derive

$$\frac{\partial \Phi_0}{\partial z} + \frac{\partial}{\partial \bar{z}} \left(\Theta_0 - 2\rho \frac{\partial W}{\partial z} \right) = 0, \quad \Phi_0 = 4\mu \frac{\partial D}{\partial \bar{z}}, \quad (4.4)$$

and from (1.3)

$$(1 - 2\sigma)\Theta_0 = 2\mu \left(\frac{\partial D_0}{\partial z} + \frac{\partial \bar{D}_0}{\partial \bar{z}} \right), \quad (1 - \sigma)(1 + \eta) = 1, \quad (4.5)$$

where by comparison with (3.2) the new elastic constant σ is called the *modified Poisson's ratio* of generalized plane stress.

Eliminating the stresses from (4.4) and (4.5) we find

$$\frac{\partial}{\partial \bar{z}} \left\{ \kappa_0 \frac{\partial D_0}{\partial z} + \frac{\partial \bar{D}_0}{\partial \bar{z}} - \frac{1}{2} \frac{\rho}{\mu} (\kappa_0 - 1) \frac{\partial W}{\partial z} \right\} = 0, \quad \kappa_0 = 3 - 4\sigma, \quad (4.6)$$

identical in form with (3.3), so that the solutions (3.5) and (3.6) hold also for the mean displacements and mean stresses of generalized plane stress, if we replace the elastic constant κ by κ_0 , i. e., provided we replace Poisson's ratio η by the modified Poisson's ratio σ .

In the sections which follow later we shall often employ generalized plane stress but drop the suffix ₀ for simplicity. The same boundary

problem can then either relate to a long cylinder in a state of plane strain or to a plate in a state of generalized plane stress, but it must be remembered that, whereas from (4.1) the faces of the plate in a state of generalized plane stress are free from surface traction, the ends of a cylinder in plane strain are under a distribution, of normal traction only, given by $\bar{z}z = \eta\Theta$.

Restriction on the Body-Force Potential.

There is also a restriction on the body-force potential in the case of generalized plane stress over and above the condition $\partial U / \partial Z = 0$. For in the case of generalized plane stress (2.3) becomes

$$B = (1 - \sigma)\Theta - 2\rho U,$$

and then (2.4) gives

$$\nabla_1^2 \{ (2 - \sigma)\Theta - 2\rho(1 + \nu)U \} + \frac{\partial^2 \Theta}{\partial Z^2} = 0, \quad \frac{\partial^2 \Theta}{\partial Z^2} = (1 + \eta)\rho\nu\nabla_1^2 U,$$

$$\left(\nabla_1^2 = 4 \frac{\partial^2}{\partial z \partial \bar{z}} \right).$$

From the latter we see that

$$\Theta = \theta_0 + \theta_1 Z + \frac{1}{2}\rho\nu(1 + \eta)Z^2\nabla_1^2 U,$$

where θ_0 and θ_1 are independent of Z , and substituting back into the former, we find that

$$\nabla_1^2 \theta_0 = (1 + \eta)\rho\nabla_1^2 U, \quad \nabla_1^2 \theta_1 = 0,$$

and

$$\nabla_1^4 U = 0. \quad \dots \dots \dots (4.7)$$

Since, if we take means of $\nabla_1^2 \Theta$ across the thickness of the plate we find

$$\nabla_1^2 \Theta_0 = \nabla_1^2 \theta_0 = (1 + \eta)\rho\nabla_1^2 U = \frac{4\rho}{\kappa_0 + 1} \nabla_1^2 U = \frac{1}{2}\rho(4 - \gamma_0)\nabla_1^2 \frac{\partial W}{\partial z},$$

and this is in agreement with the generalized plane stress solution for Θ_0 corresponding to (3.6), we see that (4.7) is the only additional restriction on the body-force potential function $U(z, \bar{z})$, or, since it must be real

$$U(z, \bar{z}) = \bar{z}U'_1(z) + z\bar{U}'_1(\bar{z}) + U'_2(z) + \bar{U}'_2(\bar{z}). \quad \dots \dots (4.8)$$

This restriction is not given in Coker and Filon's 'Photo-elasticity' ⁽⁵⁾ in their (2.24), p. 139, where they assume the harsher restriction $\nabla_1^2 \Omega = 0$ in order to simplify the differential equation for the Airy stress function, their Ω being our U , so that in effect they take $U'_2(z) = 0$. Incidentally, an error appears in their next section in their defining equation (2.252), where an extra term $\rho\Omega/(1 - \sigma)$ should be added to the right-hand member in order that their equations (2.2531) and (2.2532) should follow correctly; this leads to some consequential errors in their treatment of body-force.

The case of an elastic plate in steady rotation in its own plane about the point $z=z_0$, with angular velocity n , is given by

$$U(z, \bar{z}) = \frac{1}{2}n^2(z\bar{z}_0 + \bar{z}z_0 - z\bar{z}), \quad \dots \quad (4.9)$$

which satisfies (4.7), but makes $\nabla_1^2 U = -2n^2 \neq 0$, so that this case does not come within the scope of the Airy stress function analysis given by Coker and Filon. We give an example of this type in section 9.

Relation of the Complex Potentials and Body-Force Functions to Airy's Stress Function.

In the Airy stress function method with a stress function χ (see ⁽⁵⁾, p. 129) we have

$$\bar{x}\bar{x} = \frac{\partial^2 \chi}{\partial y^2} + \rho U, \quad \bar{y}\bar{y} = \frac{\partial^2 \chi}{\partial x^2} + \rho U, \quad \bar{x}\bar{y} = -\frac{\partial^2 \chi}{\partial x \partial y}, \quad \dots \quad (4.10)$$

$$\text{so that} \quad 2\Theta = 8 \frac{\partial^2 \chi}{\partial z \partial \bar{z}} + 4\rho U, \quad 2\Phi = -8 \frac{\partial^2 \chi}{\partial \bar{z}^2}. \quad \dots \quad (4.11)$$

Hence, from (3.6) and (4.8),

$$8 \frac{\partial^2 \chi}{\partial z \partial \bar{z}} = \Omega'(z) + \bar{\Omega}'(\bar{z}) - \gamma \rho \{z\bar{U}'_1(\bar{z}) + \bar{z}U'_1(z) + \bar{U}'_2(\bar{z}) + U'_2(z)\},$$

$$8 \frac{\partial^2 \chi}{\partial \bar{z}^2} = z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z}) - \gamma \rho \{U_1(z) + \frac{1}{2}z^2\bar{U}''_1(\bar{z}) + z\bar{U}''_2(\bar{z})\},$$

which leads to

$$\chi = \text{real part of } \frac{1}{4}\{z\bar{\Omega}(\bar{z}) + \bar{\omega}(\bar{z}) - \gamma \rho [\frac{1}{2}z^2\bar{U}_1(\bar{z}) + z\bar{U}_2(\bar{z})]\}, \quad \dots \quad (4.12)$$

showing how the Airy stress function is related to the complex potentials and the given body-force functions $U_1(z)$, $U_2(z)$.

For a complete solution, including the displacements, by the Airy stress function method, a second function, the displacement function (see ⁽⁵⁾, p. 131), has to be found, involving the solution of a further differential equation. This is often so tedious that its explicit formulation in a problem has been avoided—sometimes on the plea that the engineer is only interested in the stresses and not in the displacements—with occasionally unfortunate results, arising from the fact that it may happen that single-valued stresses are found satisfying the boundary conditions, which, however, correspond to unphysical many-valued displacements.

There is no occasion for this to happen in the complex potential method, for the same functions which give the stresses also give the displacements and their physical suitability is at once given by (3.5).

5. Transformation of Displacement and Stress.

If axes of n and s are now taken, obtained by rotating the axes of x and y through an angle α , so that

$$n + is = ze^{-i\alpha}, \quad \dots \quad (5.1)$$

then the displacement D' and the stress combinations Θ' , Φ' referred to the new axes are

$$\Theta' = \widehat{nn} + \widehat{ss}, \quad \Phi' = \widehat{nn} - \widehat{ss} + 2i\widehat{sn}, \quad D' = u_n + iu_s, \quad (5.2)$$

where $\Theta' = \Theta, \quad \Phi' = \Phi e^{-2i\alpha}, \quad D' = D e^{-i\alpha}. \quad (5.3)$

Since a mere transformation of origin without rotation of the axes gives rise to no change in the specification of stress, the new axes may conveniently be taken to be along the normal and the tangent to any curve in the x - y plane, α being the angle between the normal and the x -axis, and we have

$$\frac{\partial z}{\partial n} = e^{i\alpha}, \quad \frac{\partial z}{\partial s} = i e^{i\alpha}, \quad \frac{\partial z}{\partial s} \cdot \frac{\partial \bar{z}}{\partial s} = 1. \quad (5.4)$$

With orthogonal curvilinear co-ordinates ξ, η given by

$$z = f(\xi), \quad \zeta = \xi + i\eta, \quad (5.5)$$

we have $\frac{dz}{d\zeta} = f'(\zeta) = J e^{i\alpha}, \quad \text{or} \quad \frac{d\zeta}{dz} \frac{d\bar{z}}{d\bar{\zeta}} = e^{-2i\alpha}. \quad (5.6)$

6. Boundary Conditions in Complex Potentials.

These rules for the transformation of stress now give

$$\begin{aligned} 4(\widehat{nn} + i\widehat{ns}) &= 2(\Theta' + \Phi') = 2(\Theta + \Phi e^{-2i\alpha}) \\ \text{or} \quad 4(\widehat{nn} + i\widehat{ns}) \frac{\partial z}{\partial s} &= \left\{ \Omega'(z) + \bar{\Omega}'(\bar{z}) + (4 - \gamma)\rho \frac{\partial W}{\partial z} \right\} \frac{\partial z}{\partial s} \\ &\quad + \left\{ z\bar{\Omega}''(\bar{z}) + \bar{\omega}''(\bar{z}) - \gamma\rho \frac{\partial W}{\partial \bar{z}} \right\} \frac{\partial \bar{z}}{\partial s}, \end{aligned}$$

whence

$$4 \left\{ \widehat{nn} + i\widehat{ns} - \rho \frac{\partial W}{\partial z} \right\} \frac{\partial z}{\partial s} = \frac{\partial}{\partial s} \{ \Omega(z) + z\bar{\Omega}'(\bar{z}) + \bar{\omega}'(\bar{z}) - \gamma\rho W \}. \quad (6.1)$$

Thus, in a material under no body-force, we have, at an unstressed boundary, the simple boundary condition

$$\bar{\Omega}(\bar{z}) + \bar{z}\Omega'(z) + \omega'(z) = \text{const.} \quad (6.2)$$

(changing the sign of i for later convenience), and when there is only one boundary we can take the constant to be zero by absorbing it into one of the complex potentials.

Again for constant boundary stresses, with no body force, say an all-round pressure p and shear q , we have

$$\bar{\Omega}(\bar{z}) + \bar{z}\Omega'(z) + \omega'(z) + 4(p + iq)z = \text{const.} \quad (6.3)$$

If, instead of the stresses, the displacements are prescribed at the boundary, say $D = D_1(z, \bar{z})$, then the boundary condition is, from (3.5), after changing the sign of i throughout,

$$\kappa\bar{\Omega}(\bar{z}) - \bar{z}\Omega'(z) - \omega'(z) + \gamma\rho\bar{W} - 8\mu\bar{D}_1 = 0. \quad (6.4)$$

7. *Stress Resultants round a Circuit.*

If X , Y are the force resultants at the origin, and N the moment about the origin for the stresses round a closed circuit C , it is readily seen that

$$X+iY = \int_C (\widehat{x}\widehat{n} + i\widehat{y}\widehat{n}) ds = -i \int_C (\widehat{n}\widehat{n} + i\widehat{s}\widehat{n}) \frac{\partial z}{\partial s} ds, \quad (7.1)$$

$$N = \int_C (x\widehat{y}\widehat{n} - y\widehat{x}\widehat{n}) ds = \text{real part of } -i \int_C \bar{z}(\widehat{n}\widehat{n} + i\widehat{s}\widehat{n}) \frac{\partial z}{\partial s} ds. \quad (7.2)$$

If we take the particular body-force potential

$$U(z, \bar{z}) = cz + \bar{c}\bar{z} + d z\bar{z}, \quad (7.3)$$

which is sufficiently wide to deal with gravitational problems with gravity in the plane of the plate and making an angle β with the x -axis when

$$c = -\frac{1}{2}ge^{-i\beta}, \quad d = 0, \quad (7.4)$$

and with problems of steady rotation if, as in (4.9),

$$c = \frac{1}{2}n^2\bar{z}_0, \quad d = -\frac{1}{2}n^2, \quad (7.5)$$

then (7.1) and (7.2) give, using (6.1),

$$X+iY = -\frac{1}{4}iCy\{\Omega(z) + z\bar{\Omega}'(\bar{z}) + \bar{\omega}'(\bar{z})\} + 2\rho S(\bar{c} + dz_1), \quad (7.6)$$

$$N + iN^* = \frac{1}{4}Cy\{\bar{\omega}(\bar{z}) - \bar{z}\bar{\omega}'(\bar{z})\} - \frac{8\rho iS}{\kappa+1}\{2cz_1 + (\kappa+3)\bar{c}\bar{z}_1\}, \quad (7.7)$$

where S is the area enclosed by the circuit and $z=z_1$ gives the centroid of the area, whilst $Cy\phi$ is the notation, due to Filon, for the change in the cyclic function ϕ on passing once round the circuit C in the conventional positive sense which keeps the area enclosed by C on the left. These results are helpful in choosing tentative complex potentials to satisfy the boundary conditions.

Note also that the condition that D should be single valued is, from (3.5).

$$Cy\{\kappa\bar{\Omega}(\bar{z}) - \bar{z}\Omega'(z) - \omega'(z) + \gamma\rho\bar{W}\} = 0. \quad (7.8)$$

8. *Examples of Annular Boundary Problems.**Rigid Body Displacement of the Boundaries.*

The space between two rigid circular boundaries of radii a and b is filled with an elastic solid which adheres completely to the boundaries. The inner boundary of radius b is kept fixed, the outer is given a small rigid body displacement so that

$$D = \alpha + i\beta z \text{ over } z\bar{z} = a^2, \quad (8.1)$$

where α is a complex constant and β a real constant. We find the complex potentials which give the state of strain in the material, and

the force and couple necessary to effect the displacement, assuming no body-force. Consider the complex potentials

$$\Omega(z) = A \log z + Bz + Cz^2, \quad \dots \quad (8.2)$$

$$\omega(z) = Ez \log z + F \log z + Gz + Hz^{-1}. \quad \dots \quad (8.3)$$

For the displacement D to be single valued, (7.8) gives

$$E = -\kappa \bar{A}. \quad \dots \quad (8.4)$$

From (3.5) we find

$$8\mu D = \{ \kappa \log z \bar{z} - 2\bar{C}z\bar{z} + \kappa A - \bar{G} \} \\ + z \left\{ \kappa B - \bar{B} - \frac{\bar{F}}{z\bar{z}} \right\} + z^2 \left\{ \kappa C - \frac{\bar{A}}{z\bar{z}} + \frac{\bar{H}}{z^2\bar{z}^2} \right\},$$

after eliminating E by means of (8.4), so that $D=0$ round $z\bar{z}=b^2$ if

$$2\kappa A \log b - 2\bar{C}b^2 + \kappa A - \bar{G} = 0,$$

$$\kappa B - \bar{B} - \bar{F}b^{-2} = 0, \quad \kappa C - \bar{A}b^{-2} + \bar{H}b^{-4} = 0,$$

whilst (8.1) is satisfied round $z\bar{z}=a^2$ if

$$2\kappa A \log a - 2\bar{C}a^2 + \kappa A - \bar{G} = 8\mu\alpha,$$

$$\kappa B - \bar{B} - \bar{F}a^{-2} = 8\mu i\beta, \quad \kappa C - \bar{A}a^{-2} + \bar{H}a^{-4} = 0.$$

Solving these six equations for the constants A, B, C, F, G, H we have

$$\frac{A}{a^2+b^2} = \frac{H}{a^2b^2} = \kappa \bar{C} = \frac{\kappa \bar{G}}{\kappa^2(a^2+b^2)(2 \log b + 1) - 2b^2} \\ = \frac{4\mu\kappa\alpha}{\kappa^2(a^2+b^2) \log \frac{a}{b} - (a^2-b^2)}. \quad (8.5)$$

$$(\kappa+1)b^2B = -F = \frac{8\mu i\beta a^2b^2}{a^2-b^2}. \quad \dots \quad (8.6)$$

Hence the appropriate complex potentials for the rotational displacements are very simple, being

$$\Omega(z) = \frac{8\mu i\beta a^2z}{(\kappa+1)(a^2-b^2)}, \quad \omega(z) = -\frac{8\mu i\beta a^2b^2 \log z}{a^2-b^2}, \quad \dots \quad (8.7)$$

the rotation being effected by a couple N, given from (7.7) by

$$N = \frac{\pi}{2} iF = \frac{4\pi\mu\beta a^2b^2}{(a^2-b^2)}. \quad \dots \quad (8.8)$$

Again the translational displacement of the boundary demands a force, given from (7.6), as

$$X + iY = \frac{\pi}{2}(\kappa+1)A = \frac{2\pi\mu\kappa(\kappa+1)\alpha(a^2+b^2)}{\kappa^2(a^2+b^2) \log \frac{a}{b} - (a^2-b^2)}, \quad \dots \quad (8.9)$$

the appropriate complex potentials being

$$\Omega(z) = P\{\kappa\alpha(a^2 + b^2) \log z + \bar{\alpha}z^2\}, \quad \dots \quad (8.10)$$

$$\omega(z) = P\{-\kappa^2\bar{\alpha}(a^2 + b^2)z \log \bar{z} + \bar{\alpha}^2[\kappa^2(a^2 + b^2)(2 \log b + 1) - 2b^2]z + \kappa\alpha a^2 b^2/z\}, \quad (8.11)$$

where

$$P = \frac{4\mu}{\left[\kappa^2(a^2 + b^2) \log \frac{a}{b} - (a^2 - b^2)\right]}. \quad \dots \quad (8.12)$$

It is clear that a similar solution is easily found when the outer boundary is held fixed and the inner one given the rigid body displacement. By superposing two such solutions for equal and opposite force and couple resultants we obtain the solution for the case in which each rigid boundary is displaced slightly.

Annulus under Gravity.

We consider next the annulus $b \leq r \leq a$ of elastic material at rest under a body-force g per unit mass in the negative direction of the y -axis, the material adhering to the inner boundary which is fixed, and proceed to find the complex potentials appropriate to the two following problems:—

- (i) the outer boundary $r = a$ enclosed by and adhering to a uniform rigid cylindrical shell of weight W per unit thickness;
- (ii) the outer boundary free and unstressed.

The resultant of the stresses round the inner boundary is equivalent to a downward force equal to the weight $W + \pi\rho g(a^2 - b^2)$ of the suspended material (per unit thickness) balanced by the upward reaction of the fixings of the inner core. Hence for the stress resultants taken round the inner boundary

$$X + iY = -i\{W + \pi\rho g(a^2 - b^2)\}, \quad N = 0. \quad \dots \quad (8.13)$$

The body-force function of (7.3) and (7.4), with $\beta = -\pi/2$, makes

$$W(z, \bar{z}) = \frac{1}{2}igz\bar{z} - \frac{1}{4}igz^2, \quad \dots \quad (8.14)$$

so that the boundary condition (6.4) is

$$\kappa\bar{\Omega}'(\bar{z}) - \bar{z}\Omega'(z) - \omega'(z) - \frac{1}{2}i\gamma\rho g(z\bar{z} - \frac{1}{2}z^2) = 8\mu\bar{D}. \quad \dots \quad (8.15)$$

Now take complex potentials

$$\Omega(z) = A \log z + Bz^2, \quad \omega(z) = Cz \log z + Ez + F/z. \quad \dots \quad (8.16)$$

Then writing $z\bar{z} = r^2$, the boundary conditions for the problem (i) are

$$\begin{aligned} \{2\kappa\bar{A} \log r - 2Br^2 - C - E - \frac{1}{2}i\gamma\rho gr^2\} - (C + \kappa\bar{A}) \log z \\ + \frac{1}{z^2}\{\kappa\bar{B}r^4 - Ar^2 + F + \frac{1}{4}i\gamma\rho gr^4\} = 0 \text{ round } r = b, \\ = 8\mu i d \text{ round } r = a, \end{aligned}$$

along $z\bar{z}=a^2$, using (8.14), whence

$$\Omega(z) + z\bar{\Omega}'\left(\frac{a^2}{z}\right) + \bar{\omega}'\left(\frac{a^2}{z}\right) - \frac{1}{2}\gamma\rho ig(a^2 - \frac{1}{2}z^2) = -\rho ig(2a^2 \log z - z^2) + \text{const.}$$

along $r=a$, or

$$\{A - \bar{C} + 2i\rho ga^2\} \log z + z^2 \left\{ B + \frac{\bar{A}}{a^2} - \frac{\bar{F}}{a^4} - \frac{2\rho ig}{\kappa+1} \right\} = \text{const.}$$

along $r=a$, which is satisfied, using (8.26), if

$$B + \frac{\bar{A}}{a^2} - \frac{\bar{F}}{a^4} - \frac{2\rho ig}{\kappa+1} = 0.$$

This completes the necessary five equations for the five constants A, B, C, E, F and they are readily solved to give

$$C = -\kappa\bar{A} = -i\kappa\alpha, \quad \text{where } \alpha = 2\rho ga^2/(\kappa+1), \quad . \quad . \quad . \quad (8.27)$$

$$\bar{F} = a^4 B = \frac{1}{2}i\alpha \frac{a^2 b^2 [2a^2 + b^2(\kappa-1)]}{a^4 + \kappa b^4}, \quad . \quad . \quad . \quad . \quad (8.28)$$

$$E = 2i\kappa\alpha \log b + i\kappa\alpha \left(\frac{a^2 - b^2}{a^2} \right) + i\alpha \frac{b^2}{a^2} \frac{(a^2 - b^2)^2}{a^4 + \kappa b^4}, \quad . \quad . \quad (8.29)$$

and these equations, together with (8.16), give the complex potentials for this problem. We find the hoop stress $\widehat{\theta\theta}$ in this case round the boundary $r=a$, where $\widehat{r\bar{r}}=0$, from (3.6) and (5.3) as

$$\begin{aligned} \widehat{\theta\theta} &= \frac{1}{2} \left\{ \Omega'(z) + \bar{\Omega}'(\bar{z}) + \frac{8\rho}{\kappa+1} \frac{\partial W}{\partial z} \right\}_{r=a}, \\ &= \frac{1}{2} \left\{ \Omega'(z) + \bar{\Omega}'\left(\frac{a^2}{z}\right) + 2i\frac{\alpha}{a} \left(\frac{a}{z} - \frac{z}{a} \right) \right\}, \end{aligned}$$

$$\text{or} \quad \widehat{\theta\theta} = \frac{2\rho ga(a^2 - b^2)^2}{(\kappa+1)(a^4 + \kappa b^4)} \sin \theta, \quad . \quad . \quad . \quad . \quad . \quad (8.30)$$

using (8.27) and (8.28).

9. Elliptic Cylinder or Disk Rotating about its Axis.

As further examples of the use of the boundary condition (6.1) for an unstressed boundary, namely

$$\frac{\partial}{\partial s} \{ \Omega(z) + z\bar{\Omega}'(\bar{z}) + \bar{\omega}'(\bar{z}) - \gamma\rho W \} = -4\rho \frac{\partial W}{\partial z} \frac{\partial z}{\partial s}, \quad . \quad . \quad (9.1)$$

we consider two problems of steady rotation, treated as statical problems by the usual device of treating the reversed mass-accelerations as body-forces, given for rotation about the origin with angular velocity n , by

$$U = \frac{\partial W}{\partial z} = -\frac{1}{2}n^2 z\bar{z} \quad \text{or} \quad W = -\frac{1}{4}n^2 z^2\bar{z}. \quad . \quad . \quad . \quad (9.2)$$

For a circular disk (generalized plane stress) or circular cylinder (plane strain) given by $z\bar{z}=a^2$, (9.1) becomes

$$\frac{\partial}{\partial s}\{\Omega(z)+z\bar{\Omega}'(\bar{z})+\bar{\omega}'(\bar{z})+\frac{1}{4}\gamma\rho n^2 z^2\bar{z}\}=2\rho n^2 a^2 \frac{\partial z}{\partial s},$$

or
$$\Omega(z)+z\bar{\Omega}'(\bar{z})+\omega'(\bar{z})=2\rho n^2 a^2 z-\frac{1}{4}\gamma\rho^2 a^2 z+\text{const.}$$

along the boundary.

This is satisfied at once by the single complex potential

$$\Omega(z)=\rho n^2 a^2 \left(1-\frac{\gamma}{8}\right)z, \quad \omega(z)=0. \quad . \quad . \quad . \quad . \quad (9.3)$$

Next we take the problem of the elliptic disk (or cylinder) rotating about its centre, of boundary given by $\xi=\alpha$, where

$$z=c \cosh \zeta, \quad \zeta=\xi+i\eta, \quad . \quad . \quad . \quad . \quad (9.4)$$

round the boundary of which $\zeta+\bar{\zeta}=2\alpha$, so that

$$\bar{z}=c \cosh (2\alpha-\zeta)=c \cosh 2\alpha \cosh \zeta -c \sinh 2\alpha \sinh \zeta. \quad . \quad . \quad (9.5)$$

The boundary equation (9.1), using (9.5), becomes

$$\begin{aligned} \frac{\partial}{\partial s}\{\Omega(z)+z\bar{\Omega}'(\bar{z})+\bar{\omega}'(\bar{z})+\frac{1}{4}\gamma\rho n^2 z^2\bar{z}\} \\ =2\rho n^2 c^3 \cosh \zeta \sinh \zeta \cosh (2\alpha-\zeta) \frac{\partial \zeta}{\partial s}, \end{aligned}$$

and, on integrating and absorbing the constant into $\Omega(z)$, we have

$$\begin{aligned} \Omega(z)+z\bar{\Omega}'(\bar{z})+\bar{\omega}'(\bar{z})+\frac{1}{4}\gamma\rho n^2 z^2\bar{z} \\ =\frac{2}{3}\rho n^2 c^3 \{\cosh 2\alpha \cosh^3 \zeta -\sinh 2\alpha \sinh^3 \zeta\}, \end{aligned}$$

or, again using (9.5),

$$\begin{aligned} \Omega(z)+z\bar{\Omega}'(\bar{z})+\bar{\omega}'(\bar{z})=\frac{2}{3}\left(1-\frac{3\gamma}{8}\right)\rho n^2 c^3 (\cosh 2\alpha \cosh^3 \zeta -\sinh 2\alpha \sinh^3 \zeta) \\ +\frac{1}{4}\gamma\rho n^2 c^3 \sinh 2\alpha \sinh \zeta. \quad (9.6) \end{aligned}$$

Take the complex potentials

$$\Omega(z)=Ac^2z+Bz^3, \quad \omega(z)=Cc^2z^2+Ez^4, \quad . \quad . \quad . \quad . \quad (9.7)$$

where A, B, C, E are real constants. These make

$$\Omega(z)+z\bar{\Omega}'(\bar{z})+\bar{\omega}'(\bar{z})=2Ac^2z+2Cc^2\bar{z}+Bz^3+3Bz\bar{z}^2+4E\bar{z}^3,$$

so that along the boundary, once more using (9.5), we find

$$\begin{aligned} \Omega(z)+z\bar{\Omega}'(\bar{z})+\bar{\omega}'(\bar{z}) \\ =c^3 \cosh \zeta \{2(A+C \cosh 2\alpha)-3 \sinh^2 2\alpha(B+4E \cosh 2\alpha)\} \\ -c^3 \sinh \zeta . 2 \sinh 2\alpha \{C+3 \cosh 2\alpha(B+2E \cosh 2\alpha)\} \\ +c^3 \cosh^3 \zeta \{2B(3 \cosh^2 2\alpha-1)+4E \cosh 2\alpha(4 \cosh^2 2\alpha-3)\} \\ -c^3 \sinh^3 \zeta . 2 \sinh 2\alpha \{3B \cosh 2\alpha+2E(4 \cosh^2 2\alpha-1)\}. \quad (9.8) \end{aligned}$$

Comparison of (9.6) and (9.8) shows that we can satisfy the boundary condition by making the four expressions in brackets vanish, so that

$$3B \cosh 2\alpha + 2E(4 \cosh^2 2\alpha - 1) = \frac{1}{3}(1 - \frac{3}{8}\gamma)\rho n^2,$$

$$B(3 \cosh 2\alpha - \operatorname{sech} 2\alpha) + 2E(4 \cosh^2 2\alpha - 3) = \frac{1}{3}(1 - \frac{3}{8}\gamma)\rho n^2,$$

(whence $B + 4E \cosh 2\alpha = 0$

simplifies the remaining conditions to)

$$A + C \cosh 2\alpha = 0,$$

$$C - \frac{3}{2}B \cosh 2\alpha = 0.$$

Solving for the constants we find

$$A = -C \cosh 2\alpha = \frac{1}{8}\gamma\rho n^2 \cosh 2\alpha + (1 - \frac{3}{8}\gamma)\rho n^2 \frac{\cosh^3 2\alpha}{2 \cosh^2 2\alpha + 1}, \quad (9.9)$$

$$B = -4E \cosh 2\alpha = \frac{2}{3}(1 - \frac{3}{8}\gamma)\rho n^2 \frac{\cosh 2\alpha}{2 \cosh^2 2\alpha + 1}, \quad (9.10)$$

and then, since $a = c \cosh \alpha$, $b = c \sinh \alpha$, with the usual notation for the semi-axes of an ellipse, the complex potentials can be written

$$\Omega(z) = \frac{1}{8}\gamma\rho n^2(a^2 + b^2)z + (1 - \frac{3}{8}\gamma)\rho n^2 \frac{a^2 + b^2}{3a^4 + 2a^2b^2 + 3b^4} \left\{ \frac{2}{3}(a^2 - b^2)z^3 + (a^2 + b^2)^2 z \right\}, \quad (9.11)$$

$$\omega(z) = -\frac{1}{8}\gamma\rho n^2(a^2 - b^2)z^2 - (1 - \frac{3}{8}\gamma)\rho n^2 \frac{a^2 - b^2}{3a^4 + 2a^2b^2 + 3b^4} \left\{ \frac{1}{6}(a^2 - b^2)z^4 + (a^2 + b^2)^2 z^2 \right\}. \quad (9.12)$$

Note that when $a = b$, we have the simple results of (9.3).

It is of interest to find the hoop stress $\widehat{\eta\eta}$ on the boundary, where $\widehat{\xi\xi} = 0$, so that from (3.6) and (5.2)

$$\begin{aligned} (\widehat{\eta\eta})_{\xi=\alpha} &= \frac{1}{2} \left\{ \Omega'(z) + \overline{\Omega}'(\bar{z}) - 2 \left(1 - \frac{\gamma}{4} \right) \rho n^2 z \bar{z} \right\}_{\xi=\alpha}, \\ &= \left\{ A c^2 + \frac{3}{2} B (z^2 + \bar{z}^2) - \left(1 - \frac{\gamma}{4} \right) \rho n^2 z \bar{z} \right\}_{\xi=\alpha}. \end{aligned}$$

But the equation of the boundary can be written

$$r^2 = z\bar{z} = \frac{2a^2b^2}{a^2 + b^2} + \frac{1}{2} \frac{a^2 - b^2}{a^2 + b^2} (z^2 + \bar{z}^2),$$

and using this to eliminate $z^2 + \bar{z}^2$ we have, after some simplification,

$$(\widehat{\eta\eta})_{\xi=\alpha} = \rho n^2 \frac{(a^2 - b^2)^2 + \gamma a^2 b^2}{(a^2 - b^2)^2 + 2(a^2 + b^2)^2} (a^2 + b^2 - r^2), \quad (9.13)$$

where, in terms of Poisson's ratio η , the elastic constant γ is given by

$$\gamma = 2(1 - 2\eta)/(1 - \eta) \quad \text{or} \quad 2(1 - \eta), \quad (9.14)$$

according as we are dealing with plane strain or generalized plane stress.

10. All-Round Tension and Simple (Direct) Tension.

Consider the state of stress represented by

$$\widehat{xx}=\widehat{yy}=T, \quad \widehat{xy}=0 \quad \text{or} \quad \Theta=2T, \quad \Phi=0. \quad (10.1)$$

Changing to axes parallel to the tangent s and the normal n at a point P of a curve C , where the normal n makes an angle β with the x -axis, we have, if

$$n+is=z'=ze^{i\beta},$$

from (5.3) that

$$\Theta'=\Theta=2T \quad \text{and} \quad \Phi'=\Phi e^{-2i\beta}=0,$$

whence

$$\widehat{nn}=\widehat{ss}=T, \quad \widehat{ns}=0,$$

and this will hold for all points P round the boundary C , so that it gives a constant all-round tension T .

When there are no body-forces, the solution for the complex potentials is immediate for any shaped plate (multiply- or simply-connected, provided *all* the boundaries are acted upon by the all-round tension T) by taking

$$\Omega(z)=2Tz, \quad \omega(z)=0, \quad (10.2)$$

since this at once gives the solution (10.1), using (3.6).

A second useful elementary state of stress is given by the simple or direct tension T in a direction making an angle β with the x -axis. Here

$$\widehat{nn}=T, \quad \widehat{ss}=0, \quad \widehat{ns}=0,$$

so that

$$\Theta=\Theta'=T, \quad \Phi=\Phi'e^{-2i\beta}=Te^{-2i\beta}. \quad (10.3)$$

For a plate under such direct tension under no body-forces, the appropriate complex potentials are

$$\Omega(z)=Tz, \quad \omega(z)=-Te^{-2i\beta}z^2, \quad (10.4)$$

since from (3.6) we obtain (10.3). The appropriate stresses are supposed applied at the boundaries in this case. If the plate extends to infinity, it should be realized that in these two types of stress the displacements are unavoidably infinite at infinity.

Suppose now we have a plate extending to infinity and restrict the stresses to be finite there, *i. e.* suppose that

$$\widehat{xx} \rightarrow A, \quad \widehat{yy} \rightarrow B, \quad \widehat{xy} \rightarrow C \quad \text{as} \quad |z| \rightarrow \infty,$$

where A, B, C are finite real constants, then we can always consider the state of stress at infinity to be the superposition of an all-round tension T_1 and a simple tension T_2 in a direction β with the x -axis, such that

$$\Theta=A+B=2T_1+T_2, \quad \Phi=A-B+2iC=T_2e^{-2i\beta}, \quad (10.5)$$

since (10.5) gives A, B, C in terms of T_1, T_2 and β and *vice versa* if desired. Accordingly, in what follows we shall regard the (T_1, T_2, β) descriptions of the state of stress at infinity as fundamental, rather than the cartesian state of superposed x - and y -tensions A and B , and the simple shear C .

11. *Tension of Semi-Infinite Plate with Trochoidal Boundary.*

Suppose we transform the z -plane conformally on a ζ -plane by the relation

$$z=c(\zeta+i\lambda e^{i\xi}), \quad (11.1)$$

where c and λ are real and positive constants, so that

$$x=c(\xi-\lambda e^{-\eta} \sin \xi), \quad y=c(\eta+\lambda e^{-\eta} \cos \xi). \quad . . . (11.2)$$

Then the upper ζ -half-plane $\eta \geq 0$ corresponds to the upper half of the z -plane bounded by the trochoid $\eta=0$, as in fig. 1, whose parametric equations are

$$x=c(\xi-\lambda \sin \xi), \quad y=c\lambda \cos \xi,$$

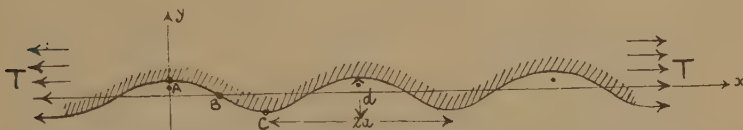
so that the curve lies between $y=\pm c$ and repeats itself after intervals of $2\pi c$, the ratio of the depth of the notches of the boundary to the half wave-length being $d/a=2\lambda/\pi$, or

$$\lambda=\pi d/2a. \quad (11.3)$$

The critical points of the transformation, given by

$$z'(\zeta)=0 \quad \text{or} \quad e^{i\xi}=1/\lambda,$$

Fig. 1.



Semi-infinite plate (under tension) with trochoidal boundary.

only lie outside the elastic material if $\lambda \leq 1$, which is also the condition that the boundary curve is sinuous and not looped, the trochoid becoming the cycloid for $\lambda=1$. We shall assume $\lambda \leq 1$, which, from (11.3) is seen to limit the depth/wave-length ratio $d/2a$, so that

$$d/2a \leq 1/\pi.$$

If there are no body-forces the single boundary $\eta=0$ will, from (6.2), be unstressed, if

$$\bar{\Omega}(\bar{z})+\bar{z}\Omega'(z)+\omega'(z)=0,$$

so that, writing

$$\Omega(z)=\Omega_0(\zeta), \quad \omega(z)=\omega_0(\zeta), \quad (11.4)$$

the boundary condition along $\eta=0$, where $\bar{\zeta}=\zeta$, is

$$z'(\zeta)\bar{\Omega}_0(\zeta)+\bar{z}(\zeta)\Omega_0'(\zeta)+\omega'(\zeta)=0. \quad (11.5)$$

Now if the plate had the straight boundary $y=0$, the complex potentials for the plate under simple tension T parallel to the x -axis are, from (10.4),

$$\Omega(z)=Tz, \quad \omega(z)=-Tz^2.$$

Accordingly, for y large, *i. e.* where η tends to infinity, we must have

$$\Omega_0(\zeta) \rightarrow Tc\zeta, \quad \omega_0(\zeta) \rightarrow -Tc^2\zeta^2,$$

and so we assume tentative complex potentials satisfying this condition,

$$\Omega_0(\zeta) = Tc(\zeta + Ae^{i\zeta}), \quad \omega_0(\zeta) = Tc^2(B\zeta - \zeta^2),$$

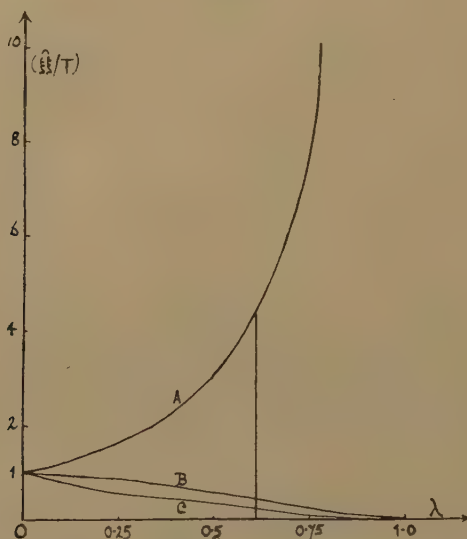
which make the boundary condition (11.5) become

$$\zeta(iA - \lambda)e^{i\zeta} + (\bar{A} - i\lambda)e^{-i\zeta} + \lambda A - \lambda\bar{A} + B = 0.$$

This is satisfied by taking

$$A = -i\lambda, \quad B = 2i\lambda^2,$$

Fig. 2.



Variation of hoop stress $\widehat{\xi\xi}$ with λ at the points A, B, C of fig. 1.

so that the complex potentials appropriate to the problem are

$$\Omega_0(\zeta) = Tc\{\zeta - i\lambda e^{i\zeta}\}, \quad \omega_0(\zeta) = Tc^2\{2i\lambda^2\zeta - \zeta^2\}. \quad (11.6)$$

We solve for the boundary hoop stress $\widehat{\xi\xi}$ given by

$$\widehat{\xi\xi} = \Theta'_{\eta=0} = \frac{1}{2}\{\Omega'(z) + \bar{\Omega}'(\bar{z})\}_{\eta=0} = \frac{1}{2}\left\{\frac{\Omega'_0(\zeta)}{z'(\zeta)} + \frac{\bar{\Omega}'_0(\bar{\zeta})}{\bar{z}'(\bar{\zeta})}\right\},$$

$$\text{whence} \quad \widehat{\xi\xi}/T = (1 - \lambda^2)/(1 + \lambda^2 - 2\lambda \cos \xi), \quad (11.7)$$

so that the stress concentration factor at the bottom of the notches is

$$(\widehat{\xi\xi}/T)_{\zeta=0} = (1 + \lambda)/(1 - \lambda). \quad (11.8)$$

The values of $(\widehat{\xi\xi}/T)$ are plotted against λ in fig. 2 for the points A, B, C of the notch [see fig. 1], and consideration of an ordinate for a given λ

makes it apparent how strongly the major variation of the hoop stress takes place in the bottom half of the notch.

Notice that the restriction $z'(\zeta) \neq 0$ within the plate shows that the solution obtained in this simple way is physically admissible. If the same solution were applied to the case where the critical points lie inside the material, we should get infinite non-admissible stresses at these points. The cusp-form cracks of the cycloid boundary inevitably lead to an infinite hoop stress at these points.

Wave-like boundaries similar to the above of more general form have recently been treated by Weber⁽⁶⁾ by the Airy stress function method. The above serves to emphasize a point not always brought out in extensive formal work on general boundaries, that there must be implied restrictions on the coefficients in the defining equations of the boundary, if the solution found by just such tentative means as above is to hold good. The solution for the periodic boundary, which consists of the joined arcs of the looped trochoidal boundary but omitting the loops themselves, for example, cannot be as simple in form.

12. Infinite Plate with Curvilinear Polygonal Hole.

Consider the transformation

$$z = c\sigma(1 + \lambda\sigma^{-n}), \quad \sigma = e^{\xi}, \quad \zeta = \xi + i\eta, \quad \text{where } 0 \leq \lambda(n-1) \leq 1, \quad (12.1)$$

c, λ being real positive constants and n a positive integer. Then

$$|\sigma| \leq 1 \quad \text{when } z'(\sigma) = c[1 - \lambda(n-1)\sigma^{-n}] = 0,$$

and so the critical points of the transformation lie within the unit circle in the σ -plane and within its transform in the z -plane. The subsidiary transformation $\sigma = e^{\xi}$ merely gives us useful curvilinear co-ordinates ξ, η such that the boundary of the hole in the elastic z -plane corresponds to the unit circle $\sigma\bar{\sigma} = 1$ in the σ -plane, *i. e.* is given by $\xi = 0$.

Putting $z = re^{i\theta}$, the parametric equations of the boundary of the hole are

$$r^2 = z\bar{z} = c^2(1 + \lambda^2 + 2\lambda \cos n\eta), \quad \tan \theta = \frac{\sin \eta - \lambda \sin (n-1)\eta}{\cos \eta + \lambda \cos (n-1)\eta}, \quad (12.2)$$

so that r is stationary when $\sin n\eta = 0$, or $\eta = s\pi/n$, $s = 0, 1, 2, 3 \dots$. From (12.1) and (12.2), the curve is such that

$$\frac{n-2}{n-1} \leq 1 - \lambda \leq r \leq 1 + \lambda \leq \frac{n}{n-1}. \quad \dots \quad (12.3)$$

Also, when $\sin n\eta = 0$, $\tan \theta = \tan \eta$, so that θ and η increase together. It is clear that the boundary is accordingly a certain curvilinear regular polygon of n "sides" and n rounded "vertices," these occurring where r is a maximum, *i. e.* $\cos n\eta = +1$, and becoming cusps when $\lambda = 1/(n-1)$ when the critical points of the conformal transformation fall on the boundary of the unit circle. We shall suppose that $n \geq 2$, since $n = 1$ gives a circular hole and the transformation to the σ -plane is superfluous.

For $n=2$ the hole is elliptic, the semi-axes a and b in the usual notation being given by

$$a=c(1+\lambda), \quad b=c(1-\lambda), \quad . \quad . \quad . \quad . \quad (12.4)$$

and the transformation is then the well-known elliptic transformation. For ξ large both $|\sigma|$ and $|z|$ are large, so the distant parts of the z - and ξ -planes correspond.

The curvature of the boundary $\xi=0$ follows readily from the formula ^(3), 1942)

$$\frac{\partial \alpha}{\partial s} + i \frac{\partial \alpha}{\partial n} = z''(\sigma) \bar{z}'(\bar{\sigma}) / \{z'(\sigma) \bar{z}'(\bar{\sigma})\}^{3/2},$$

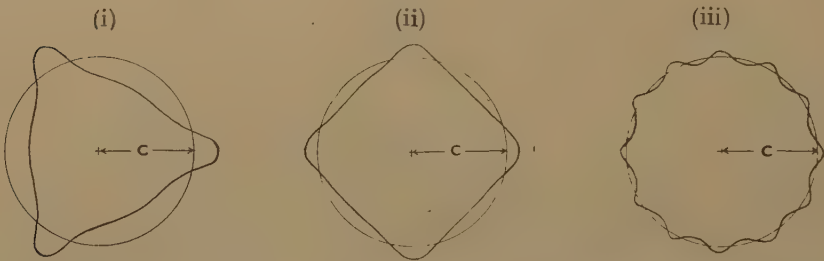
leading to

$$\frac{\partial \alpha}{\partial s} = \frac{1 + \lambda(n-1)^2 \cos n\eta - \lambda(n-1) \cos n\eta - \lambda^2(n-1)^3}{c\{1 - 2\lambda(n-1) \cos n\eta + \lambda^2(n-1)^2\}^{3/2}},$$

so that at the points of greatest r ($\cos n\eta = +1$) the radius of curvature ρ is given by

$$\frac{\rho}{c} = \frac{\{1 - \lambda(n-1)\}^2}{1 + \lambda(n-1)^2}, \quad . \quad . \quad . \quad . \quad (12.5)$$

Fig. 3. -



$$(i) \quad n=3, \lambda = \frac{1}{4} \left(< \frac{1}{2} \right). \quad (ii) \quad n=4, \lambda = \frac{1}{9} \left(< \frac{1}{3} \right). \quad (iii) \quad n=12, \lambda = \frac{1}{16} \left(< \frac{1}{11} \right).$$

Examples of curvilinear polygonal boundary curve of hole.

whereas at the points of least r ($\cos n\eta = -1$) the radius of curvature is given by

$$\frac{\rho}{c} = \frac{\{1 + \lambda(n-1)\}^2}{1 - \lambda(n-1)^2}, \quad . \quad . \quad . \quad . \quad (12.6)$$

and we see that for $\lambda = 1/(n-1)^2$ (which satisfies the restricting inequality of (12.1)), the polygons are "flat" at the mid-points of their sides. Fig. 3 shows the form of these particular flat-sided polygons when $n=3$ and when $n=4$, the remaining boundary is that for $n=12$, $\lambda = 1/16$.

Unstressed Hole in Infinite Plate under All-round Tension.

In the case of no body-forces, with the notation of (11.4), the boundary condition (6.2) for the single boundary $\sigma\bar{\sigma}=1$ becomes

$$z'(\sigma)\bar{\Omega}_0(\sigma^{-1}) + \bar{z}(\sigma^{-1})\Omega_0'(\sigma) + \omega_0'(\sigma) = 0. \quad . \quad . \quad . \quad . \quad (12.7)$$

Take the complex potentials

$$\Omega_0(\sigma) = 2Tc\sigma + Ac\sigma^{-(n-1)}, \quad \omega_0(\sigma) = Cc^2 \log \sigma, \quad . \quad . \quad (12.8)$$

where A and C are real constants. From (10.1) and (12.1) it is readily seen that these lead to an all-round stress at infinity of amount T . On dividing out c^2 , (12.7) becomes

$$\{4T + C - 2A\lambda(n-1)\}\sigma^{-1} + (A + 2T\lambda)\{\sigma^{n-1} - (n-1)\sigma^{-(n+1)}\} = 0,$$

which is satisfied by taking

$$A = -2T\lambda, \quad C = -4T[1 + \lambda^2(n-1)],$$

so that the appropriate complex potentials for the problem are

$$\Omega_0(\sigma) = 2cT\{\sigma - \lambda\sigma^{-(n-1)}\}, \quad \omega_0(\sigma) = -4c^2T[1 + \lambda^2(n-1)] \log \sigma, \quad (12.9)$$

for, although $\omega_0(\sigma)$ is a cyclic function, it is at once apparent from (7.7) and (7.8) that it gives rise to zero-stress resultants round the hole and to single-valued stresses and displacements.

The hoop stress $\widehat{\eta\eta}$ round the hole is given by

$$\widehat{\xi\xi} + \widehat{\eta\eta} = \frac{1}{2}\{\Omega'(z) + \bar{\Omega}'(\bar{z})\} = \frac{1}{2}\left\{\frac{\Omega'_0(\sigma)}{z'(\sigma)} + \frac{\bar{\Omega}'_0(\bar{\sigma})}{\bar{z}'(\bar{\sigma})}\right\},$$

whence

$$(\widehat{\eta\eta}/T)_{\xi=0} = 2[1 - (n-1)^2\lambda^2]/\{1 - 2\lambda(n-1)\cos n\eta + (n-1)^2\lambda^2\}, \quad (12.10)$$

so that at the vertices of the polygon, where $\cos n\eta = +1$, the stress concentration factor is

$$\left(\frac{\widehat{\eta\eta}}{T}\right)_{\xi=0} = \frac{2[1 + (n-1)\lambda]}{[1 - (n-1)\lambda]}, \quad \dots \dots \dots (12.11)$$

whilst at the mid-points of the sides it has the reciprocal value.

If we let n tend to infinity, which from (12.1) entails that λ tends to zero, (12.10) gives the well-known result 2 for this factor at all points of a circular hole. In contrast the values given by (12.11) for the three boundaries of fig. 3 are respectively 6, 4 and 10.8.

When $n=2$ (12.10) has a simple geometrical form, for if ϖ is the perpendicular from the origin to the tangent to the boundary

$$\begin{aligned} \varpi &= r \sin \phi = \text{coefficient of } i \text{ in } re^{i(\frac{\pi}{2} + \alpha - \theta)} \\ &= \text{real part of } \bar{z} \sqrt{\frac{dz}{d\xi} \frac{d\bar{z}}{d\bar{\xi}}}, \text{ i. e. of } \frac{\bar{z}(\bar{\sigma})\sigma z'(\sigma)}{\sqrt{\bar{z}'(\bar{\sigma})z'(\sigma)}} \\ &= c(1 - \lambda^2)/\{1 - 2\lambda \cos \widehat{n\eta} + \lambda^2\}, \text{ when } n=2, \end{aligned}$$

hence (12.10) gives

$$(\widehat{\eta\eta}/T)_{\xi=0} = 2T\varpi^2/ab. \quad \dots \dots \dots (12.12)$$

Unstressed Hole in Infinite Plate under Simple Tension at Infinity.

Now take the complex potentials

$$\left. \begin{aligned} \Omega_0(\sigma) &= c\{T\sigma + A\sigma^{-1} + B\sigma^{-(n-3)} + C\sigma^{-(n-1)}\}, \\ \omega_0(\sigma) &= c^2\{-Te^{-2i\beta}\sigma^2 + F \log \sigma + G\sigma^{-2} + H\sigma^{-(n-2)}\}, \end{aligned} \right\} \quad (12.13)$$

where F is a real constant, so that, from (7.7) there is no couple resultant about the hole, but the constants A, B, C, G, H may be complex. Since as $|\sigma| \rightarrow \infty$, $c\sigma \rightarrow z$, we see from (10.4) that the complex potentials give a stress system at infinity which tends to that of simple tension T at an angle β with the x -axis, whilst it is clear from (7.8) that the displacements are single-valued. The boundary condition (12.7) becomes

$$\begin{aligned} & \{2T + F - \lambda(n-1)(C + \bar{C})\}\sigma^{-1} - \{A + \lambda(n-1)\bar{B} + 2G\}\sigma^{-3} \\ & + \{\bar{C} + \lambda T\}\sigma^{n-1} - (n-1)\{C + \lambda T\}\sigma^{-(n+1)} + \{\bar{B} - A\lambda\}\sigma^{n-3} \\ & + \{\bar{A} - 2Te^{-2i\beta} - \lambda(n-3)B\}\sigma - \{(n-3)B + (n-2)H + \lambda(n-1)\bar{A}\}\sigma^{-(n-1)} = 0, \end{aligned}$$

which is satisfied by making the various coefficients of the powers of σ vanish separately, giving

$$\begin{aligned} C = \bar{C} &= -\lambda T, \quad F = 2T[1 + \lambda^2(n-1)], \\ H = -2B = -2\lambda\bar{A} &= -\frac{4\lambda Te^{-2i\beta}}{1 - \lambda^2(n-3)}, \quad G = -Te^{-2i\beta} \frac{1 + \lambda^2(n-1)}{1 - \lambda^2(n-3)}. \end{aligned}$$

Hence the appropriate complex potentials for the problem are

$$\Omega_0(\sigma) = Tc\{\sigma - \lambda\sigma^{-(n-1)}\} + 2Tc \frac{e^{2i\beta}\sigma^{-1} + \lambda e^{-2i\beta}\sigma^{-(n-3)}}{1 - \lambda^2(n-3)}, \quad (12.14)$$

$$\begin{aligned} \omega_0(\sigma) &= Tc^2\{-e^{-2i\beta}\sigma^2 + 2[1 + \lambda^2(n-1)\log\sigma] \\ &\quad - Tc^2 \frac{\{e^{2i\beta}[1 + \lambda^2(n-1)]\sigma^{-2} + 4\lambda e^{-2i\beta}\sigma^{-(n-2)}\}}{1 - \lambda^2(n-3)}\}. \end{aligned} \quad (12.15)$$

Interest in this solution centres on the stress concentration factor at the corners of the hole. Writing $\Omega_0(\sigma) = Tc\Sigma A_r\sigma^r$ for brevity, we have

$$\left(\frac{\eta\eta}{T}\right)_{\sigma=1} = \frac{1}{2T} \left\{ \frac{\Omega_0'(\sigma)}{z'(\sigma)} + \frac{\bar{\Omega}_0'(\sigma^{-1})}{\bar{z}'(\sigma^{-1})} \right\}_{\sigma=1} = -\frac{\Sigma r(A_r + \bar{A}_r)}{2[1 - \lambda(n-1)]},$$

or

$$\left(\frac{\eta\eta}{T}\right)_{\sigma=1} = \frac{1}{1 - \lambda(n-1)} \left\{ 1 + \lambda(n-1) - \frac{2 \cos 2\beta [1 + \lambda(n-3)]}{1 - \lambda^3(n-3)} \right\}, \quad (12.16)$$

gives the stress concentration factor at the particular vertex $\sigma=1$ or $\zeta=0$; its value at the remaining vertices can be deduced by replacing β by $\beta - 2r\pi/n$, $r=1, 2, \dots, n-1$. This extends the result for $\beta=\pi/2$ communicated to the writer by H. L. Cox, to whom indeed the present writer owes the idea of applying the complex potential method to this particular curvilinear boundary, a very useful one for exhibiting the power of the complex potential method.

For $\beta=\pi/2$, i. e. the plate is under tension in the y -direction

$$\left(\frac{\eta\eta}{T}\right)_{\sigma=1} = \frac{1}{1 - \lambda(n-1)} \left\{ 1 + \lambda(n-1) + \frac{2[1 + \lambda(n-3)]}{1 - \lambda^2(n-3)} \right\}. \quad (12.17)$$

As $n \rightarrow \infty$ (so $\lambda \rightarrow 0$) this tends to the well-known value 3, whilst in contrast for the three boundaries of fig. 3 under these conditions, the values are respectively 7, 5.375, 15.764.

Curvilinear Polygonal Hole under Uniform Boundary Stresses.

We suppose the plate to be under stresses which tend to vanish at infinity, whilst the hole is under a constant pressure p and constant shear stress q , so that from (6.3), writing $p + iq = \alpha$ for brevity,

$$\bar{z}(\sigma^{-1})\Omega_0'(\sigma) + \omega_0'(\sigma) + z'(\sigma)\{\bar{\Omega}_0(\sigma^{-1}) + 4\alpha\bar{z}(\sigma^{-1})\} = 0. \quad (12.18)$$

Take $\Omega_0(\sigma) = A\sigma^{-(n-1)}, \quad \omega_0(\sigma) = C\sigma^2 \log \sigma + B\sigma^{-n}, \quad (12.19)$

which are readily seen to lead to vanishing stresses at infinity, then, on dividing out c^2 , (12.18) becomes

$$\{C - \lambda(n-1)(A + \bar{A}) + 4\alpha[1 - \lambda^2(n-1)]\}\sigma^{-1} + \{\bar{A} + 4\alpha\lambda\}\sigma^{n-1} - \{nB + A(n-1) + 4\alpha\lambda(n-1)\}\sigma^{-(n+1)} = 0,$$

which is satisfied by taking

$$A = -4\lambda\bar{\alpha}, \quad C = -4\alpha - 4\bar{\alpha}(n-1)\lambda^2, \quad B = 4\lambda \frac{(n-1)}{n}(\bar{\alpha} - \alpha),$$

so that the appropriate complex potentials are

$$\Omega_0(\sigma) = -4\bar{\alpha}\lambda\sigma^{-(n-1)}, \quad (12.20)$$

$$\omega_0(\sigma) = -4c^2[\alpha + \alpha(n-1)\lambda^2] \log \sigma + 4\lambda \frac{(n-1)}{n} c^2(\bar{\alpha} - \alpha)\sigma^{-n}. \quad (12.21)$$

For the hoop stress round the boundary we have

$$\widehat{\eta\eta} = \frac{1}{2} \left\{ \frac{\Omega_0'(\sigma)}{z'(\sigma)} + \frac{\bar{\Omega}_0'(\sigma^{-1})}{\bar{z}'(\sigma^{-1})} \right\}_{\xi=0},$$

$$\widehat{\eta\eta} = \frac{4\lambda(n-1)\{p \cos n\eta + q \sin n\eta - p\lambda(n-1)\}}{1 - 2\lambda(n-1) \cos n\eta + \lambda^2(n-1)^2}. \quad (12.22)$$

For the ellipse $n=2$, this gives

$$\widehat{\eta\eta} = 2p^2 \left(\frac{\varpi^2}{ab} - 1 \right) + 2q \frac{xy \varpi^2}{a^3 b^3}, \quad (12.23)$$

where ϖ is the perpendicular from the centre on the tangent.

For the polygonal hole under constant pressure p , the stress concentration factor at the corners of the hole is

$$\left(\frac{\widehat{\eta\eta}}{p} \right) = \frac{4\lambda(n-1)}{1 - \lambda(n-1)}, \quad (12.24)$$

which, for the boundaries pictured in fig. 3, has the values respectively 4, 2, 0.55. For the circle the factor is zero.

For the polygonal hole under constant shear q , the couple applied to the boundary is equal and opposite to the couple N about any circuit reducible to the boundary of the hole, which, from (7.7) is given by

$$N = -2\pi c^2 q [1 - \lambda^2(n-1)]. \quad (12.25)$$

Small Rotation of Rigid Curvilinear Core.

Imagine an infinite elastic medium under no body-force, adhering to a rigid core embedded in it, whose cross-section is the curvilinear boundary $\sigma\sigma=1$, and which is given a small rigid body rotation through an angle β about the origin, so that the boundary displacement is

$$D = i\beta z.$$

Then, from (6.4), the boundary condition round $\xi=0$ can be written

$$\kappa z'(\sigma)\bar{\Omega}(\sigma^{-1}) - \bar{z}(\sigma^{-1})\Omega'(\sigma) - \omega'(\sigma) + 8\mu i\beta z(\sigma^{-1})z'(\sigma) = 0. \quad (12.26)$$

Take the complex potentials

$$\Omega_0(\sigma) = A\sigma^{-n-1}, \quad \omega_0(\sigma) = B\sigma^2 \log \sigma + C\sigma^2\sigma^{-n}, \quad (12.27)$$

which lead to one-valued stresses and displacements tending to vanish at infinity, and then (12.26) can be written as

$$[\kappa\bar{A} + 8\mu i\beta\lambda]\sigma^{n-1} + [nC + (n-1)A - 8\mu i\beta\lambda(n-1)]\sigma^{-(n+1)} \\ + \{8\mu i\beta[1 - \lambda^2(n-1)] + (A - \kappa\bar{A})\lambda(n-1) - B\}\sigma^{-1} = 0,$$

which can be satisfied by taking

$$A = 8\mu i\beta \frac{\lambda}{\kappa}, \quad B = 8\mu i\beta \left[1 + \frac{\lambda^2}{\kappa}(n-1)\right], \quad C = 8\mu i\beta\lambda \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{\kappa}\right).$$

Hence the appropriate complex potentials are

$$\Omega_0(\sigma) = 8\mu i\beta \frac{\lambda}{\kappa} \sigma^{-(n-1)}, \quad \omega_0(\sigma) = 8\mu i\beta c^2 \left[1 + \frac{\lambda^2}{\kappa}(n-1)\right] \log \sigma \\ + 8\mu i\beta\lambda c^2 \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{\kappa}\right) \sigma^{-n}. \quad (12.28)$$

Interest here centres on the couple to be applied to the rigid core, which is $-N$, where, from (7.7), the couple N about any circuit reducible to the hole is

$$N = -4\pi\mu\beta c^2 \left[1 + \frac{\lambda^2}{\kappa}(n-1)\right]. \quad (12.29)$$

Small Displacement of Translation of Rigid Curvilinear Core.

When the boundary displacement is $D=\alpha$, where α is a small complex constant, the boundary condition (6.4) becomes

$$\kappa z'(\sigma)\bar{\Omega}(\sigma^{-1}) - \bar{z}(\sigma^{-1})\Omega'(\sigma) - \omega'(\sigma) - 8\mu\bar{\alpha}z'(\sigma) = 0. \quad (12.30)$$

Take complex potentials

$$\Omega_0(\sigma) = A \log \sigma + B\sigma^{-(n-2)}, \quad \omega_0(\sigma) = Cz \log \sigma + Fc\sigma^{-1} + Gc\sigma^{-(n-1)}, \quad (12.31)$$

which makes the boundary condition

$$-z'(\sigma) \log \sigma \{\kappa\bar{A} + C\} + c\sigma^{n-2} \{\kappa\bar{B} - \lambda A\} - c\sigma^{-2} \{\lambda(n-1)\kappa\bar{B} + A - F\} \\ + c\sigma^{-n} \{(n-2)B - \lambda C + (n-1)G + 8\mu\bar{\alpha}\lambda(n-1)\} + c\{(n-2)\lambda B - C - 8\mu\bar{\alpha}\} = 0,$$

This is satisfied if all the expressions in the $\{ \}$ brackets vanish separately, leading to

$$\frac{F}{1 + \lambda^2(n-1)} = \kappa \frac{\bar{B}}{\lambda} = -\frac{\bar{C}}{\kappa} = A = \frac{8\mu\alpha\kappa}{\kappa^2 + \lambda^2(n-2)} \quad (12.32)$$

$$-G = 8\mu\bar{\alpha}\lambda \left\{1 + \frac{\kappa^2 + (n-2)}{(n-1)[\kappa^2 + \lambda^2(n-2)]}\right\}. \quad (12.33)$$

Hence the complex potentials are determined. The solution is more appropriate to a finite plate with appropriate stresses applied at the outer boundary, since the displacements contain an (unavoidably) infinite term $\kappa A \log \sigma\bar{\sigma}$ at infinity. The stresses at infinity do, however, tend to vanish, but they have a force resultant $X+iY$, given by (7.6), as

$$X+iY = \frac{4\mu\pi\alpha\kappa(\kappa+1)}{\kappa^2 + \lambda^2(n-2)}, \quad (12.34)$$

and the force which has to be applied to the rigid core to effect this displacement is $-(X+iY)$.

13. Circular Disk under Given Edge Stresses.

General series solutions for single circular boundary problems are very easily constructed.

Suppose the disk $r=a$ is in equilibrium under the given edge loading

$$(\widehat{rr}+i\widehat{r\theta})_{r=a}=f(\theta)=\sum_{-\infty}^{+\infty}\alpha_n e^{in\theta}, \quad \dots \quad (13.1)$$

where α_n is complex and given by

$$\alpha_n = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) e^{-in\theta} d\theta. \quad \dots \quad (13.2)$$

Since the disk is in equilibrium, the stress resultants $X+iY$ and N round the boundary must vanish, and since

$$X+iY = \int_0^{2\pi} a(\widehat{rr}+i\widehat{r\theta})e^{i\theta} d\theta, \quad N = \int_0^{2\pi} a^2 \widehat{r\theta} d\theta, \quad \dots \quad (13.3)$$

(13.1) leads to

$$\alpha_{-1}=0, \quad \alpha_0=\bar{\alpha}_0 \text{ or } \alpha_0 \text{ is real.} \quad \dots \quad (13.4)$$

Now take the complex potentials

$$\Omega(z) = a \sum_0^{\infty} A_n \left(\frac{z}{a}\right)^{n+1}, \quad \omega(z) = a^2 \sum_2^{\infty} B_n \left(\frac{z}{a}\right)^n, \quad \dots \quad (13.5)$$

since $\log z$ and negative powers of z would lead to infinite stresses at the origin. The constants are complex, although we shall take A_0 as real here, since an imaginary A_0 , like the terms $(B_0+B_1z/a)a^2$ omitted from $\omega(z)$, gives rise to no stresses, and so corresponds to rigid body displacements only.

In polar co-ordinates

$$z=re^{i\theta}, \quad e^{-2i\theta}=\bar{z}/z$$

and

$$4(\widehat{rr}+i\widehat{r\theta})=2(\Theta'+\Phi')=2(\Theta+e^{-2i\theta}\Phi),$$

so that, changing the sign of i throughout and using (3.6),

$$4(\widehat{rr}-i\widehat{r\theta})_{r=a}=\Omega'(z)+\bar{\Omega}'\left(\frac{a^2}{z}\right)-z\Omega''(z)-\frac{z^2}{a^2}\omega''(z). \quad \dots \quad (13.6)$$

Then from (13.1) and (13.5)

$$4 \sum_{-\infty}^{+\infty} \bar{\alpha}_n e^{-in\theta} = \sum_0^{\infty} (n+1) \bar{A}_n e^{-in\theta} - \sum_0^{\infty} (n-1) \{nB_n + (n+1)A_n\} e^{in\theta},$$

which is satisfied by taking

$$A_0=2\alpha_0, \quad A_n=\frac{4\alpha_n}{n+1}, \quad n=1, 2, 3 \dots,$$

$$B_n=-\frac{4}{n(n-1)} \{(n-1)\alpha_n+\alpha_{-n}\}, \quad n=2, 3, 4 \dots \quad (13.7)$$

Disk under Normal Pressures over Equal and Opposite Arcs.

As an example suppose that

$$\widehat{r\theta}=0 \text{ all round } r=a, \quad \widehat{rr}=-p \text{ for } -\alpha<\theta<\alpha, \quad \pi-\alpha<\theta<\pi+\alpha. \quad (13.8)$$

Then (13.2) gives

$$\alpha_{2n}=-\frac{p}{n\pi} \sin 2n\alpha, \quad \alpha_0=-2p \frac{\alpha}{\pi}, \quad \alpha_{2n+1}=0, \quad \dots \quad (13.9)$$

so that from (13.7)

$$\begin{aligned} A_{2n+1}=0, \quad B_{2n+1}=0, \quad A_0=-4p \frac{\alpha}{\pi}, \quad A_{2n}=-\frac{4p \sin 2n\alpha}{\pi n(2n+1)}, \\ B_{2n}=\frac{4p \sin 2n\alpha}{\pi n(2n-1)}. \quad \dots \quad (13.10) \end{aligned}$$

The hoop stress $\widehat{\theta\theta}$ round the boundary is given by

$$(\widehat{rr}+\widehat{\theta\theta})_{r=a}=\frac{1}{2}\{\Omega'(z)+\bar{\Omega}'(\bar{z})\}_{r=a}=\sum_0^{\infty}(2n+1)A_{2n} \cos 2n\theta,$$

$$(\widehat{rr}+\widehat{\theta\theta})_{r=a}=-4p \frac{\alpha}{\pi}-4p \sum_1^{\infty} \frac{1}{n} \sin 2n\alpha \cos 2n\theta.$$

For $\theta=0$ and π this becomes

$$\widehat{\theta\theta}=p \left(1-4 \frac{\alpha}{\pi}\right)-4p \sum_1^{\infty} \frac{1}{n} \sin 2n\alpha=p \left(1-4 \frac{\alpha}{\pi}\right)-p \left(2-4 \frac{\alpha}{\pi}\right)=-p, \quad (13.11)$$

and for $\theta=\frac{1}{2}\pi$ and $\frac{3}{2}\pi$

$$\widehat{\theta\theta}=-4p \frac{\alpha}{\pi}-4p \sum_1^{\infty} (-1)^n \frac{\sin 2n\alpha}{n}=-4p \frac{\alpha}{\pi}+8p \frac{\alpha}{\pi}=4p \frac{\alpha}{\pi}, \quad (13.12)$$

using the Fourier expansions

$$2 \sum_1^{\infty} (-1)^{n-1} \frac{\sin nx}{n}=x, \quad -\pi<x<\pi; \quad 2 \sum_1^{\infty} \frac{\sin nx}{n}=\pi-x, \quad 0<x<2\pi. \quad \dots \quad (13.13)$$

Note that as $\alpha \rightarrow \frac{1}{2}\pi$ (13.11) gives the correct hoop stress for the disk under all-round tension $-p$, as given by section 10.

14. Circular Hole in Infinite Plate.

Suppose the edge stresses round the hole are given by (13.1) and (13.2). Then, if the stress resultants for the applied edge stresses round the hole are $-(X+iY)$ and $-N$, (13.3) leads to

$$X+iY=2\pi a\alpha_{-1}, \quad N=i\pi a^2(\bar{\alpha}_0-\alpha_0). \quad \dots \quad (14.1)$$

Take the complex potentials

$$\Omega(z)=A_0 a \log z + A_1 z + a \sum_2^{\infty} A_n \left(\frac{z}{a}\right)^{-n+1}, \quad \dots \quad (14.2)$$

$$\omega(z)=B_0 a z \log z + C_0 a^2 \log z + C_1 z^2 + a^2 \sum_1^{\infty} B_n \left(\frac{z}{a}\right)^{-n}, \quad \dots \quad (14.3)$$

where A_1 will be taken real, and A_1 and C_1 will be taken as given constants, since, from section 10, they determine the state of stress at infinity.

From (7.6) and (7.7) we find

$$X + iY = \frac{1}{2}\pi a(A_0 - \bar{B}_0), \quad N = \frac{1}{4}i\pi a^2(C_0 - \bar{C}_0). \quad (14.4)$$

Again, if there are no dislocations in the elastic material, the condition (7.8) for single valued displacements gives

$$\bar{B}_0 + \kappa A_0 = 0. \quad (14.5)$$

From (14.1), (14.4) and (14.5), we then have

$$A_0 = 4\alpha_{-1}/(\kappa + 1), \quad C_0 - \bar{C}_0 = 4(\bar{\alpha}_0 - \alpha_0). \quad (14.6)$$

The complex potentials (14.2) and (14.3) make the boundary condition (13.6)

$$\begin{aligned} 4(\widehat{rr} - i\widehat{r\theta}) = & \left\{ A_0 \frac{a}{z} + A_1 - \sum_2^\infty (n-1) A_n \left(\frac{z}{a} \right)^{-n} \right\} \\ & + \left\{ \bar{A}_0 \frac{z}{a} + A_1 - \sum_2^\infty (n-1) \bar{A}_n \left(\frac{z}{a} \right)^n \right\} \\ & - \left\{ -A_0 \frac{a}{z} + \sum_2^\infty n(n-1) A_n \left(\frac{z}{a} \right)^{-n} \right\} \\ & - \left\{ B_0 \frac{z}{a} - C_0 + 2C_1 \frac{z^2}{a^2} + \sum_1^\infty n(n+1) B_n \left(\frac{z}{a} \right)^{-n} \right\}, \end{aligned}$$

or, using (13.1)

$$\begin{aligned} 4 \sum_{-\infty}^\infty \bar{\alpha}_n e^{-in\theta} = & (2A_1 + C_0) + 2(A_0 - B_1)e^{-i\theta} + (\bar{A}_0 - B_0)e^{i\theta} - (\bar{A}_2 + 2C_1)e^{2i\theta} \\ & - \sum_3^\infty (n-1) \bar{A}_n e^{in\theta} - \sum_2^\infty (n+1) \{(n-1)A_n + nB_n\} e^{-in\theta}, \end{aligned}$$

which is satisfied by equating coefficients of $e^{in\theta}$, $n=0, \pm 1, \pm 2 \dots$, giving equations consistent with (14.5) and (14.6) and completing with these equations the coefficients in the complex potentials in terms of the boundary stress coefficients α_n and the coefficients A_1 and C_1 , which give the state of stress at infinity, as

$$C_0 = 4\bar{\alpha}_0 - 2A_1, \quad B_1 = -\frac{4\alpha_{-1}}{\kappa + 1} - 2\bar{\alpha}_1, \quad A_2 = -4\alpha_{-2} - 2\bar{C}_1. \quad (14.7)$$

$$A_n = -\frac{4\alpha_{-n}}{n-1}, \quad n=3, 4 \dots; \quad B_n = \frac{4}{n} \left\{ \alpha_{-n} - \frac{\alpha_n}{n+1} \right\}, \quad n=2, 3 \dots \quad (14.8)$$

This problem was dealt with—making some use of the complex variable, but in a rather different manner—by Bickley⁽⁷⁾.

Hole under Normal Pressures over Equal and Opposite Arcs.

As a first example, consider the problem of the perforated plate under stresses vanishing at infinity, and the stress system round the boundary

of the hole given by (13.8) and (13.9), which lead, from (14.7) and (14.8) to

$$A_1 = C_1 = 0, \quad A_0 = B_0 = 0, \quad B_{2n+1} = 0, \quad A_{2n+1} = 0, \quad \dots \quad (14.9)$$

$$A_{2n} = \frac{4p \sin 2n\alpha}{\pi n(2n-1)}, \quad B_{2n} = -\frac{4p \sin 2n\alpha}{\pi n(2n+1)}, \quad C_0 = -8p \frac{\alpha}{\pi}. \quad (14.10)$$

The hoop stress $\widehat{\theta\theta}$ round the hole is given by

$$\begin{aligned} (\widehat{rr} + \widehat{\theta\theta})_{r=a} &= \frac{1}{2} \left\{ \Omega'(z) + \overline{\Omega}'\left(\frac{a^2}{z}\right) \right\} = -\sum_1^{\infty} (2n-1) A_{2n} \cos 2n\theta \\ &= -4 \frac{p}{\pi} \sum_1^{\infty} \frac{1}{n} \sin 2n\alpha \cos 2n\theta. \quad \dots \quad (14.11) \end{aligned}$$

For $\theta=0$ and π , using (13.13), this gives

$$\widehat{\theta\theta} = p - 4 \frac{p}{\pi} \sum_1^{\infty} \frac{\sin 2n\alpha}{n} = p \left(4 \frac{\alpha}{\pi} - 1 \right), \quad \dots \quad (14.12)$$

and for $\theta = \frac{1}{2}\pi$ and $\frac{3}{2}\pi$ it gives

$$\widehat{\theta\theta} = -4 \frac{p}{\pi} \sum_1^{\infty} (-1)^n \frac{\sin 2n\alpha}{n} = 8p \frac{\alpha}{\pi}. \quad \dots \quad (14.13)$$

Hole under Normal Pressures over an Arc.

As a second example, which involves a stress resultant about a hole, consider the boundary stress over $r=a$, given by

$$\widehat{r\theta} = 0 \text{ everywhere,} \quad \widehat{rr} = -p \text{ over } |\theta| < \alpha, \quad \widehat{rr} = 0 \text{ elsewhere.} \quad (14.14)$$

Then (13.2) gives

$$\alpha_0 = -p \frac{\alpha}{\pi}, \quad \alpha_n = -\frac{p}{\pi} \frac{\sin n\alpha}{n} = \alpha_{-n}, \quad \dots \quad (14.15)$$

and, if the stresses are to vanish at infinity, (14.5)–(14.8) give

$$\begin{aligned} A_1 = C_1 = 0, \quad A_0 = -\frac{B_0}{\kappa} = -\frac{4p \sin \alpha}{\pi(\kappa+1)}, \quad C_0 = -4p \frac{\alpha}{\pi}, \quad B_1 = -\frac{2p(\kappa-1)}{\pi(\kappa+1)}, \\ \dots \quad (14.16) \end{aligned}$$

$$A_n = \frac{4p \sin n\alpha}{\pi n(n-1)}, \quad B_n = -\frac{4p \sin n\alpha}{\pi n(n+1)}, \quad n=2, 3, 4, \dots, \quad (14.17)$$

From (14.1) the stress resultant $X+iY = -2\pi ap \sin \alpha$, so that the forces applied to the hole have a resultant

$$R = 2ap \sin \alpha \quad \dots \quad (14.18)$$

along the x -axis, balanced by the evanescent stresses at infinity.

The hoop stress round the hole is given by

$$(\widehat{rr} + \widehat{\theta\theta})_{r=a} = \frac{1}{2} \left\{ \Omega'(z) + \overline{\Omega}'\left(\frac{a^2}{z}\right) \right\} = A_0 \cos \theta - \sum_2^{\infty} (n-1) A_n \cos n\theta,$$

since A_n is here real, whence

$$(\widehat{rr} + \widehat{\theta\theta})_{r=a} = 4 \frac{p}{\pi} \frac{\kappa}{\kappa+1} \sin \alpha \cos \theta - 4 \frac{p}{\pi} \sum_1^{\infty} \frac{1}{n} \sin n\alpha \cos n\theta. \quad (14.19)$$

For $\theta=0, \frac{1}{2}\pi, \pi$ respectively, using (13.13) and (4.5) and (4.6), this gives

$$\widehat{\theta\theta}=p\left(2\frac{\alpha}{\pi}-1\right)+(3-\eta)\frac{p}{\pi}\sin\alpha, \quad . \quad . \quad . \quad (14.20)$$

$$\widehat{\theta\theta}=2p\frac{\alpha}{\pi}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14.21)$$

$$, \quad \widehat{\theta\theta}=2p\frac{\alpha}{\pi}-(3-\eta)\frac{p}{\pi}\sin\alpha, \quad . \quad . \quad . \quad . \quad . \quad . \quad (14.22)$$

where η is Poisson's ratio, results which agree with Bickley's values for this problem.

Conclusion.

This and previous investigations⁽³⁾ of the present writer arose from the desire to reduce the differences of method in teaching classical two-dimensional elasticity and hydrodynamics. The gain arising from a thoroughgoing use of the complex variable—both in the approach to the problem and in its solution—is undeniable. Compare, for example, the problem represented by (12.12) and (12.13) with $n=2$ for the elliptic hole, with its usual treatment by the Airy stress function method.

The novelty of this approach to the problem is not as recent as the writer had previously supposed⁽³⁾, 1940) and dates back to 1908, when Kolossov⁽⁴⁾ began a series of papers including a joint paper with Muschelisvili in 1915. Further papers by Muschelisvili and others have since appeared^{(8), (9), (10), (11)} of considerable interest and importance, revealing both the power and the elegance of the complex variable methods.

The reader is referred to Muschelisvili⁽⁸⁾, 1932) for the proofs of the uniqueness theorems on which the uniqueness of the solutions of the problems by the tentative methods of the present paper rests.

In view of progress made so long ago, it is remarkable that no clue to a really effective use of the complex variable appears in the treatises on the subject.

References.

- (1) Filon, Proc. Lond. Math. Soc. xliii. p. 152 (1937).
- (2) Seth, Proc. Ind. Acad. Sci. v. a, p. 518 (1937).
- (3) Stevenson, Proc. Roy. Soc. (as yet unpublished) (1940); Phil. Mag. xxxiii. p. 639 (1942); xxxiv. p. 105 (1943).
- (4) Kolossov, *Comptes Rendus*, cxlvi. p. 522 (1908); cxlviii. p. 1242 (1909); *Z. f. Math. u. Phys.* lxii. p. 384 (1914); *Comptes Rendus*, clxxxi. p. 24 (1925); clxxxiv. p. 512 (1927); exciii. p. 383 (1931).
Kolossov and Muschelisvili, *Annales de l'Inst. Electrotechnique à Petrograd*, xii. p. 39 (1915).
- (5) Coker and Filon, 'Photo-Elasticity,' C.U.P. (1931).
- (6) Weber, *Z.A.M.M.* xxii. p. 29 (1942).
- (7) Bickley, Phil. Trans. Roy. Soc. ccxxvii. p. 383 (1928).
- (8) Muschelisvili, *Bull. de l'Acad. des Sc. de Russie*, xiii. p. 663 (1919); *Math. Annalen*, cvii. p. 282 (1932); *Z.A.M.M.* xiii. p. 264 (1933); *C.R. d'Acad. des Sc. de l'U.S.S.R.* iii. pp. 7, 73, 141 (1934).
- (9) Malkin, *Z.A.M.M.* x. p. 182 (1930).
- (10) Scherman, *C. R. (Doklady) Acad. Sci. U.R.S.S.* xxvi. p. 627; xxvii. pp. 329, 911; xxviii. pp. 25, 28 (1940).
- (11) Gatewood, Phil. Mag. xxxii. p. 282 (1941).

LXXXVIII. *Notices respecting New Books.*

Mathematical Tables and Aids to Computation. By R. A. ARCHIBALD. [I. No. 1, Jan. 1943 (pp. 31) ; I. No. 2, April 1943 (Pp. 37).] (National Research Council, 2101 Constitution Avenue, Washington, D.C. Price £1 per annum.)

THERE are many ways of classifying mathematical tables ; among others, we might distinguish between those which merely facilitate work and those without which it could not be done. Tables of x^2 or of $1/x$, of which Barlow's is a well-known example, or of e^x , fall into the former class, because we could, if necessary, compute them by logarithms whenever these functions are needed. Of the other class, tables of the gamma function or of the probability integral are familiar examples. A statistician requiring values of the probability integral would have to stop his work if no tables were available.

Tables of the "unessential" class may be just as valuable as those of the other, as anyone who has needed $(1-x^2)^{\frac{1}{2}}$ for many values of x will agree. Thus, whichever type of table is needed, a great deal of effort will be wasted if there is no means of finding where it can be obtained. For this reason the writer, and no doubt many readers, began some years ago to keep a note book in which mathematical tables were indexed.

Notebooks prepared by individuals are, however, liable not to be kept up-to-date, and in any case, they cannot hope to be complete. Now the National Research Council of the U.S.A. has come to our rescue by issuing a quarterly journal, at the remarkably low subscription of £1 per annum, to be known familiarly as M.T.A.C., of which the first two issues are now under review.

The Editor of the journal is R. A. Archibald, the Chairman of the Committee on Mathematical Tables. The first issue contains notices of 88 tables published in the last ten years, 14 of them being reviewed in detail. The list is very varied and includes the Smithsonian Physical Tables and Percival's 'Mathematical Facts and Formulæ.' Then follow notices of errata found in published tables, including more than 150 last-figure errors in Gifford's 'Natural Sines,' and some in Briggs's work of 1633 and Euler's of 1764. Many readers may like to know that in the 4th Edition of Barlow, the last two figures in the cube root of 197 are transposed, and that in the 3rd edition the difference following $\sqrt{10n}$ for 1156 should be 46494 and not 45494.

Another valuable section gives details of the contents and location of the MSS. of unpublished tables. Finally, there is a "Queries and Replies" section, which up to the present deals only with the question "What is the use of multi-figured tables?"

The second issue has for its main contents a long and interesting essay on trigonometrical tables in non-sexagesimal arguments (including time). This is followed by the sections on new tables, errata in published tables, and queries and replies. The journal will deal not only with tables, but also with mechanical aids to computation, but up to the present this seems to be a minor item in the contents list.

There is no doubt about the value of the journal, which should be in the library of every institution where mathematical calculation is carried out. The index is to be of such a nature that tables on particular subjects can easily be traced, even after the lapse of years.

J. H. A.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

LXXXIX. *The Photographic Determination of Flame Temperatures in Closed-Vessel Explosions.*

By A. SMEETON LEAH, Ph.D., A.M.I.Mech.E.*

[Received August 25, 1943.]

[Plate IV.]

THE temperature of flame gases is usually determined by methods involving contact of the gases with surface. This may be the surface of particles or other bodies introduced into the flame gases as, for example, in the case of resistance thermometers, thermo-couples and the spectral line-reversal method, or it may be contact with the surface of a calorimeter. These methods assume that the surfaces introduced have no effect upon the thermal and chemical equilibrium of the flame gases, and so yield values of the true temperature corresponding to the mean molecular translational energy of the molecules. Owing to the existence of latent energy and abnormal dissociation in flame gases, it is distinctly improbable that such is the case. In fact, David † has shown that for this reason the sodium line-reversal method yields incorrect temperatures. It has also been shown ‡ that platinum resistance thermometers similarly disturb the quasi-equilibrium obtaining in the abnormally dissociated products of the carbon-monoxide flame and yield far too high temperatures owing to recombination proceeding on the platinum surface.

In view of these considerations it will be obvious that there is a very real need of a method of determining flame temperatures which will be quite independent of surface effects.

No very convenient and direct methods have yet been evolved, but of the indirect methods, flame photography offers one solution, and here two lines of approach present themselves.

In the first method, as used by Flock and Roeder §, the inflammable mixture is contained in a soap bubble and after ignition the flame temperature is inferred from the photographically determined expansion-ratio. Alternatively, the temperatures can be found by determining the rise of pressure in a closed-vessel explosion and correlating this with the photographically determined volume of the inflamed gases. It is this second method which has been adopted in the following work.

* Communicated by W. T. David, Sc.D.

† Phil. Mag. ser. 7, vol. xxiii. p. 251 (1937); 'Engineering,' vol. cxxxviii. p. 475 (1934).

‡ David and Pugh, 'Nature,' vol. cxlv. p. 896 (1940); David, Leah and Pugh, Phil. Mag. ser. 7, vol. xxxi. p. 156 (1941).

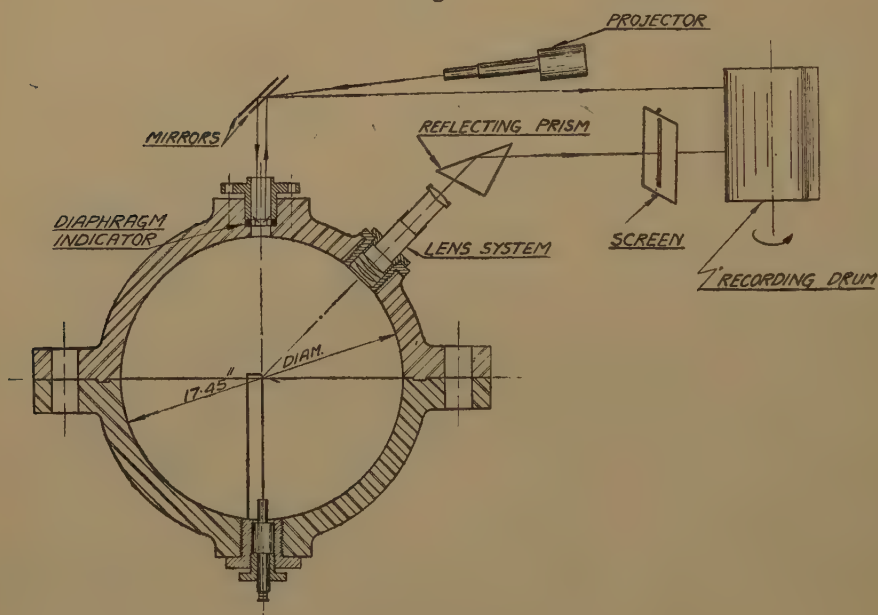
§ Nat. Adv. Comm. Aer. Rep. No. 532 (1935).

One advantage of carrying out the measurements in a closed-vessel explosion is that the variation in temperature during combustion of successive layers of gas can be studied. By this means it has been possible, for the first time, to demonstrate experimentally the validity of David's postulate *, namely that "the latent energy in closed-vessel explosions decreases with distance of flame travel from the igniting source."

Experimental.

A spherical steel explosion vessel, 17.45 inches in internal diameter, was equipped with spark electrodes giving central ignition and also with a very sensitive diaphragm type of pressure indicator. The

Fig. 1.



indicator diaphragm was made of invar to minimize temperature effects, and in addition, a small corrugation was machined in the rim of the diaphragm for the same reason. In other respects the indicator was very similar to that of Thorp †, except that instead of having only one small mirror fixed to the diaphragm, two mirrors were fitted at 90° to each other. By this means the sensitivity of the indicator was increased, as described by Nakanishi and Kitamura ‡.

An optical system was also fitted into a hole in the wall of the sphere in order to photograph the flame movement. This system consisted of three plano-concave lenses, each of 3 inches focal length, placed end to

* "Abnormality in Flame Gases," Proc. Inst. Mech. Eng. (in press).

† Phil. Mag. vol. viii. p. 813 (1929).

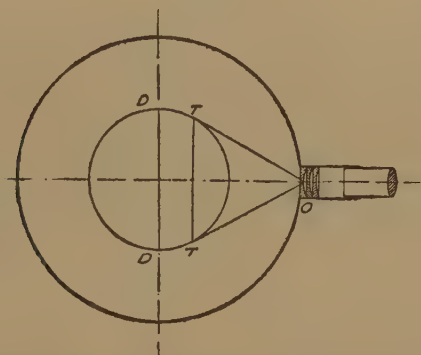
‡ Rep. Aer. Res. Inst., Tokyo, vol. vii, no. 87 (Sept. 1934).

end, the inner one being flush with the interior wall of the sphere. A convex lens was fitted at the end of a telescopic tube, by means of which the image of the flame-front could be focused on to the recording drum. A screen, perforated with a narrow slit, was interposed between the lens and the recording drum so that the photographic record would be confined to the flame movement across a diameter of the sphere. The general arrangement of the apparatus will be seen from the diagrammatic layout given in fig. 1.

Unfortunately, the maximum flame diameter which could be recorded, owing to the natural limitation of the angle of vision of the lenses, was only 10 inches. The whole history of the explosion could not therefore be recorded.

The gases used during the work, namely carbon monoxide and oxygen, were manufactured by the British Oxygen Company and stored in steel

Fig. 2.



cylinders. They were carefully analyzed for purity and the mixtures were invariably checked by measuring the contraction in volume after combustion. In certain preliminary experiments it was found that the carbon-monoxide became contaminated with iron-carbonyl on standing in the steel cylinders. The iron-carbonyl appears to have a profound effect upon the combustion of carbon monoxide. The carbon monoxide was therefore passed through a filter containing active carbon, which effectively eliminated the impurity.

Calibration of the Diaphragm Indicator and Lens System.

Calibration of the pressure indicator was effected by applying a known pressure dynamically with compressed air and photographically recording the rise on the film. Check calibrations were made by applying statical pressure to the diaphragm. In all cases the statical calibration was about 1 per cent. below the dynamic.

The lens combination had to be calibrated in order to determine the width of image recorded on the film by a luminous ball of flame of a given diameter centrally placed in the explosion vessel. It will be seen from the diagram in fig. 2 that, since the inner plano-concave lens is flush

with the vessel-wall, the diameter DD will not be the outermost image on the recording drum. Instead, if lines OT are drawn from the optical centre of the lenses tangential to the sphere, these will correspond to the outermost fringes of the image on the drum. In view of this difficulty, and because of the impracticability of having luminous spheres inside the vessel to calibrate directly, two alternative methods were used.

In the first method a brass tube of $\frac{3}{4}$ -inch diameter, having slots cut in its wall at 1-inch intervals, was placed across a diameter of the sphere at right-angles to the axis of the lenses. The slits were illuminated by means of 6 volt electric lamps placed inside the brass tube. The images of these slits were photographed on the drum, giving a series of lines on the film whose distances apart could be accurately measured by means of a micrometer. From these measurements it was possible to calculate the correction to be applied for the sphericity of the flame-front. The corrections were made by applying the theory of thick lenses and determining the size of image created by the edge of the flame-front giving the widest angle of illumination, namely TT in fig. 2.

The second method was virtually a direct calibration of the lenses. Cylinders of thick white drawing paper of different diameters were made by fastening the paper round the inside of wooden rings bored out to the exact diameters required. These rings were illuminated by means of an electric lamp placed inside. The lens system was then removed from the explosion vessel intact, without altering the focusing, and was mounted in the correct position relative to one of the rings, namely with its inner lens face 8.725 inches from the centre of the ring. A screen was then adjusted so as to be in focus and the size of image of a narrow section of the cylinder was measured. By using cylinders of different diameters the whole range of sizes could be dealt with. The calibrations obtained by the two methods agreed within 0.25 per cent.

From the photographic records of flame-travel and pressure-rise it is possible to find the pressure-rise within the sphere when the flame-front during explosion has reached any given diameter within the limits of the lenses. With the present lens system it was found impossible to record accurately diameters in excess of 10 inches, and for diameters less than 6 inches the pressure-rise was too small to be measured accurately by the diaphragm indicator. Hence the measurements here described are confined to the flame-travel from 3 to 5 inches from the spark out of a total travel of 8.725 inches.

Assuming that the unburnt gas ahead of the flame-front is being adiabatically compressed by the expanding flame, and this assumption cannot be greatly in error in view of the rapid rate of compression and the small temperature range, it is easy to show that the mean temperature of the burnt gases is given by the expression:—

$$T = \frac{T_0 r \rho}{e} \left[\frac{1}{1 - r^{\gamma-1}(1 - \rho)} \right],$$

where T_0 = initial temperature of the mixture,

r = the pressure ratio at the instant when the flame has a given diameter D ,

$$\rho = \left(\frac{D}{17.45} \right)^3,$$

$$e = \text{Ratio} \frac{\text{number of molecules after combustion}}{\text{number of molecules before combustion}},$$

$\gamma = C_p/C_v$ = ratio of specific heats,

= 1.4 for the small range of temperature considered

The greatest source of error is in correlating the rise of pressure with the diameter of the flame, although at the smallest diameters the rise of pressure is in itself quite small and is therefore difficult to measure accurately. The possible errors in temperature would appear to be of the order of $\pm 100^\circ \text{C.}$ at the smallest flame diameter of 6 inches and $\pm 60^\circ \text{C.}$ at a diameter of 10 inches.

Temperature Variation with Flame-Travel.

The inflammable mixtures chosen for the work consisted of carbon monoxide and oxygen with large excess of carbon-monoxide. In such mixtures dissociation is small and the complication of making laborious corrections is largely eliminated. Furthermore, since it has to be assumed that the flame-front travels outwards in concentric circles (the diameter only is being measured and from this the volume of burnt gas is calculated), the mixtures, must not be too slow-burning or convective rise will distort the spherical flame-front, as shown by Ellis*.

On the other hand, the burning must not be too rapid or it will be difficult to correlate pressure-rise with flame-spread. The mixtures chosen have about the required speed of burning except in the case of the weakest mixtures, when the addition of a little hydrogen is necessary to speed up the combustion.

The range of mixtures which can be conveniently dealt with is from 9 per cent. O_2 + 91 per cent. CO to 14 per cent. O_2 + 86 per cent. CO, although in the present work the attention is largely confined to the explosion of a 13 per cent. O_2 + 87 per cent. CO mixture. The whole range of experiments, together with a comparison with other methods of temperature determination, will be reported in a subsequent paper.

The reproduction of a photographic record for a typical explosion is shown in Pl. IV. The pressure record is given by the line marked P, and the instant the spark passed is marked by the spots S_1 and S_2 . The times from ignition must be read from the two different spots, S_1 for the pressure-record and S_2 for the flame-movement, since, for experimental reasons, these two spots are not aligned laterally on the record.

The results of a typical experiment with a 13 per cent. O_2 + 87 per cent. CO mixture are given in the table. The initial pressure of the explosion

* 'Fuel,' vol. vii. p. 195 (1928).

was 700 mm. Hg and the initial temperature 20° C. These results are plotted in fig. 3 on a base of time from the moment of inflammation. The trace of the flame-front is shown together with the pressure-ratio. Below these is given the curve of temperatures taken from the last column of the table.

Fig. 3.

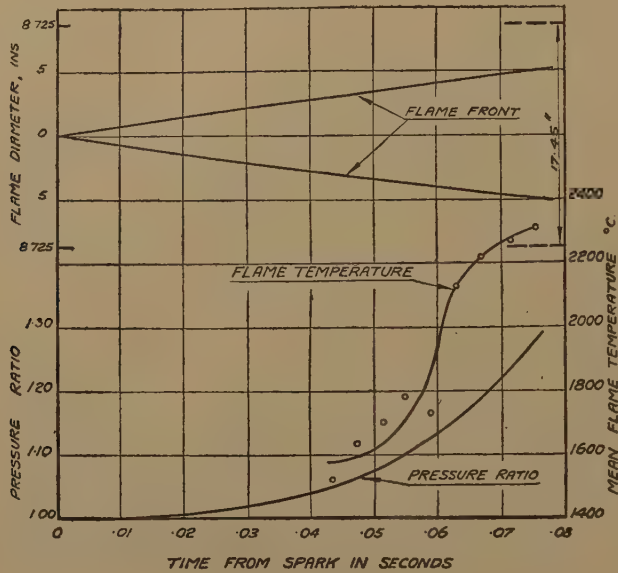


TABLE.

Diameter of flame in inches.	Time from spark in seconds.	Pressure ratio <i>r</i> .	Mean temperature of flame gases °C.
6	0.0435	1.048	1520
6½	0.0475	1.062	1630
7	0.0515	1.080	1700
7½	0.0550	1.100	1780
8	0.0590	1.125	1730
8½	0.0630	1.156	2130
9	0.0670	1.190	2220
9½	0.0715	1.233	2270
10	0.0755	1.281	2310

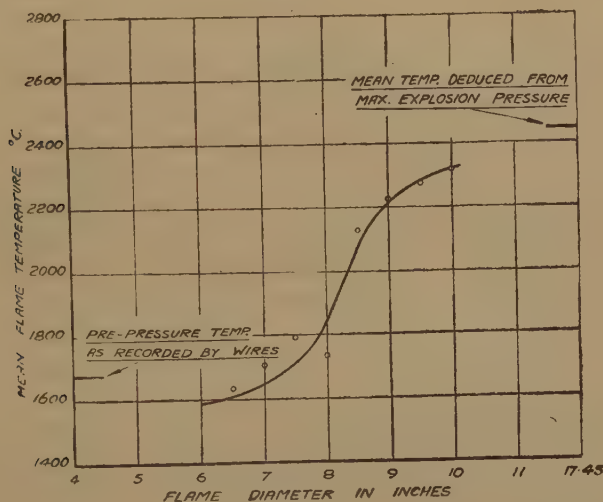
It will be evident that, when using a very sensitive indicator, there is no real “pre-pressure” period, although the time during which the flame is growing to about 6 inches diameter (in an 18 inch sphere) is generally taken to be that period. During this initial stage of combustion it has been shown by David* that the gas temperature, as determined by

* Phil. Mag. ser. 7, vol. xvii. p. 172 (1934) ; vol. xxiii. p. 345 (1937).

exceptionally fine platinum-rhodium resistance thermometers, is far below the ideally calculated value. This shows the existence of a large latent energy in the flame gases. On the other hand, pressure and heat loss measurements in closed-vessel explosions * show that at the instant of maximum pressure the latent energy is, in general, much smaller than during the "pre-pressure" period.

These observations led David † to postulate that during combustion of gaseous mixtures in closed vessels the latent energy decreases with distance of flame-travel from the igniting source. Furthermore, since it has been shown that turbulence has no effect on the initial stages of combustion, the suggestion was made that the combustion might be propagated by chain reaction mechanism. Later on, the pressure in the

Fig. 4.



flame builds up, and this combined with the heat transmission ahead of the flame-front is largely responsible for the propagation.

It will be obvious that the results of the experiment given in the table are absolutely in line with these views. To give a more complete picture, fig. 4 has been drawn showing the mean gas temperatures taken from the last column of the table plotted against flame diameter. The "pre-pressure" platinum resistance thermometer temperature ‡ is shown and the final mean gas temperature deduced from the maximum explosion

* David, Brown, and El Din. *Phil. Mag.* ser. 7, vol. xiv. p. 764 (1932); David and Leah, *Phil Mag.* ser. 7, vol. xxii. p. 513 (1936).

† *Phil. Mag.* ser. 7, vol. xxiii. p. 345 (1937); also *Proc. Inst. Mech. Engrs.* "The Abnormaity of Flame Gases" (in Press).

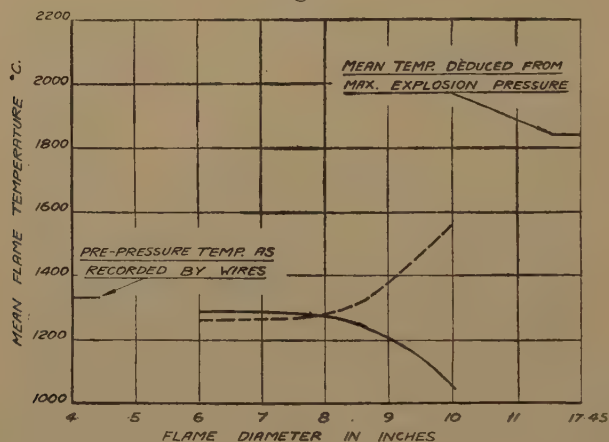
‡ It has been shown (*Phil. Mag.* ser. 7, vol. xxxi. p. 156 (1941)) that a plain platinum resistance thermometer records the same temperature as a quartz-coated platinum resistance thermometer in the case of very over-rich mixtures.

pressure is indicated. The mean temperature will be seen to increase with flame diameter at such a rate that at a diameter of 10 inches the value is not far removed from the overall mean temperature obtained from the maximum pressure*.

This shows that as the flame moves outwards from 6 inches diameter the latent energy decreases from the high value expected from platinum thermometry to the low value deduced from the maximum explosion temperature.

In examining this curve it has to be borne in mind that the temperatures are mean values for each diameter of flame. It will be seen that the rate of rise of mean temperature is high, and therefore the instantaneous temperature in the flame-front must undergo a very rapid increase shortly after the flame reaches a diameter of 6 inches. The experimental errors do not allow of any very close analysis of the curve, but it might

Fig. 5.



be stated that there is even the possibility that a transient type of latent energy exists in the very early stages of combustion and that the gases quickly lose this surplus energy, causing the mean temperature to rise at a very great rate.

In the case of a 9 per cent. O_2 + 91 per cent. CO mixture the explosion is rather too slow to prevent convective rise of the flame gases and the diameter, and therefore the volume, of the flame gases is underestimated. The result of this is to give temperatures which are too low, as will be seen by inspecting the full line curve of fig. 5. When, however, $\frac{1}{2}$ per cent. of hydrogen is added to the mixture, the explosion is sufficiently speeded up to prevent convection and the results are again in line with the stronger mixtures. This is shown by the broken-line curve of fig. 5.

In fig. 5 the platinum resistance thermometer temperature and the

* It may be noted that, since the pressure rise is small, the rise of temperature due to adiabatic compression of the gas behind the flame-front is small.

overall explosion temperature are also indicated. It will again be seen that the result of the speeded-up explosion is absolutely in line with the results of the previously described experiment, giving further confirmation to David's views.

Summary.

A new method of determining flame temperatures of carbon monoxide mixtures has been developed. The method depends upon flame photography and is, therefore, quite independent of surface contact with the flame gases.

The temperatures obtained in closed-vessel explosions show an increasing mean flame temperature and therefore a diminishing latent energy, with distance of flame-travel from the igniting source, thus confirming David's views.

XC. Electromagnet Polepieces and Condensers with Special Shapes.

By N. DAVY, University College, Nottingham *.

[Received July 12, 1943.]

IN molecular beam researches, particularly of the deflexion type, and also in magnetic susceptibility measurements, magnetic or electric fields with special properties are often needed. In order to discuss certain questions, let the field strength, whether electric or magnetic, be called H . Two types of cases occur, viz., those in which it is desirable to have (a) $\frac{\partial H^2}{\partial s}$ constant, or (b) $\frac{\partial H}{\partial s}$ constant, over a certain length of beam or throughout a small volume, where s is a coordinate of distance. Thus, when induction occurs, so that the molecules in a beam under test acquire an induced electric or magnetic moment $\frac{\partial H^2}{\partial s}$ is a factor in the expression for the lateral deflexion of the beam. When no induction occurs and the molecules have a fixed electric or magnetic moment, $\frac{\partial H}{\partial s}$ is a factor in the expression for the lateral deflexion. In susceptibility measurements it is advantageous to have either $\frac{\partial H^2}{\partial s}$ or $\frac{\partial H}{\partial s}$ constant over the small volume occupied by a specimen. H itself must sometimes be large, sometimes small, and the field may be either electric or magnetic.

* Communicated by Professor L. F. Bates, Ph.D., D.Sc.

The commonest experimental arrangement hitherto used is the Gerlach-Stern system of "cylindrical" electromagnet polepieces in the form of a wedge facing a slot in a plane (see fig. 1 A (a)). The field H and its differential coefficients in the gap between the poles are obtained by a point-to-point experimental survey. No theoretical description of such a field seems to have been worked out. This type is popular because it

gives a large value of $\frac{\partial H}{\partial s}$ at a certain point between the slot and the wedge. The molecular beam is usually sent through the gap parallel to a generator of the cylindrical system, so that except for the edge effects at entry and exit $\frac{\partial H}{\partial s}$ is practically constant along the beam.

For example, Meissner and Scheffers (1933) used Gerlach-Stern polepieces

9.9 cm. long. They found that $\frac{\partial H}{\partial s}$ only varied by about ± 2.5 per cent.

from a mean value of about 2.48×10^4 oersteds/cm., over a length of

6.70 cm. of beam. $\frac{\partial H}{\partial s}$ is not constant in directions perpendicular to

the undeflected beam. In particular, it increases rapidly in the lateral direction approaching the wedge. This causes a peculiar cusp-like excrescence (Abrutschen) to appear in the figure obtained by photographic development of the film deposited by the deflected beam on a glass plate. This want of symmetry is particularly undesirable.

Other things being equal, it would appear advantageous to use fields in which values of H , $\frac{\partial H}{\partial s}$ and $\frac{\partial H^2}{\partial s}$ can be predicted by theory at every point, so that the theory can be tested by experiment, the two lending mutual support. This applies particularly to those systems of condenser plates or polepieces which theory predicts should have $\frac{\partial H}{\partial s}$ or $\frac{\partial H^2}{\partial s}$ constant at every point in the gap between them. The purpose of the present note is to report progress in devising and testing such fields, both electric and magnetic.

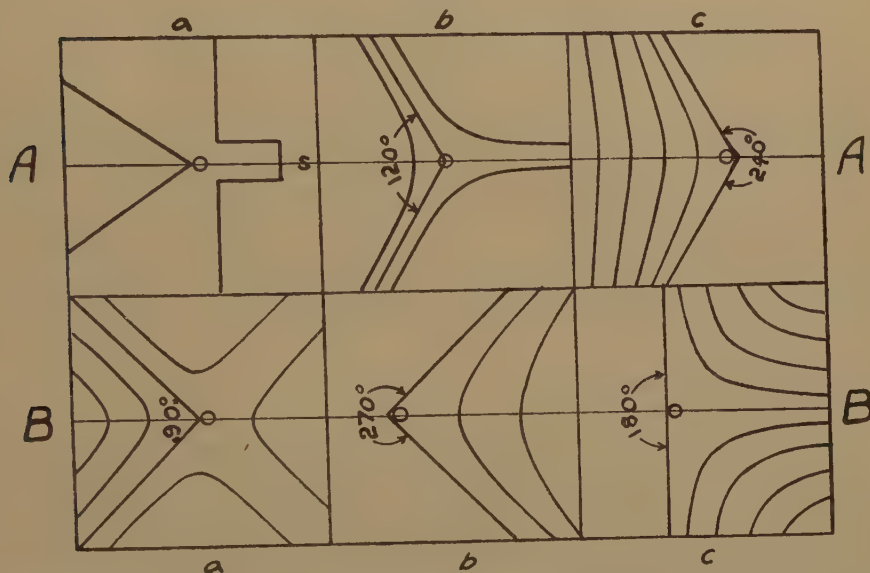
(1) *Electrical Field with $\frac{\partial H^2}{\partial s}$ constant.*

E. McMillan (1931) obtained an electric field in which $\frac{\partial H^2}{\partial s}$ was large and constant by the use of a three-plate system of cylindrical electrical conductors ⁽¹⁾. The polar equation of the curves of normal section of these conductors was $r^{3/2} \cos \frac{3\theta}{2} = \text{constant}$. McMillan used this system to deflect a beam of HCl molecules, but his published account was very short and without numerical data. The three plates had sections of

the shape represented in fig. 1 A (b) and were at three different potentials.

It may be pointed out that the arrangement of any two of the cylindrical condenser plates whose sections are represented in fig. 1 A (c) also belongs to the system $r^{3/2} \cos \frac{3\theta}{2} = \text{constant}$, and gives $\frac{\partial H^2}{\partial s}$ a constant value at every point in the gap except near the top and bottom edges. H is small near O . The capacity per unit length is $\frac{1}{2\pi} \tan \frac{3\theta}{2}$ e.s.u. per cm., where θ = the angle subtended by a half-plate at O ⁽²⁾. It would

Fig. 1.



seem possible to make such a condenser with its gap narrow so as to make $\frac{\partial H^2}{\partial s}$ large as well as constant, and thus produce suitable deflexions of molecular beams.

It would perhaps be advantageous to use defining slits whose lateral edges are concentric circles whose centre is O and whose ends are portions of radii, for the force on a molecule of constant electric moment μ_e , though constant in magnitude is radial in direction. A beam of molecules with different velocities, and with a cross-section defined by such slits, would be drawn out radially on passing through the above field.

(2) Magnetic Field with $\frac{\partial H^2}{\partial s}$ constant.

E. M. Somekh, working in Professor Bates' laboratory at University College, Nottingham, is at present (December, 1942) measuring magnetic

susceptibilities, using an electromagnet with cylindrical polepieces, the cross-sections of whose faces have the polar equation $r^{3/2} \cos \frac{3\theta}{2} = \text{constant}^{(2)}$.

By experiment he finds $\frac{\partial H^2}{\partial s}$ to be constant in value to ± 2.5 per cent.

over a wide range of points along the axis of the gap, as theory predicts. The shapes used are of the type represented in fig. 1 A (c), so that the field H is small near O. Polepieces of type fig. 1 A (b) have not yet been made, nevertheless a comparison between them and the Stern-Gerlach type is perhaps worth making, by the aid of figs. 1 A (a) and A (b). Both types are cylindrical.

The Stern-Gerlach shape, fig. 1 A (a), gives high values of $\frac{\partial H}{\partial s}$ in the gap, constant in the longitudinal direction, varying in the lateral direction. The same applies to $\frac{\partial H^2}{\partial s}$. The wedge has an angle of the order 70° . This faces a slotted plane. The polepieces defined by the equation $r^{3/2} \cos \frac{3\theta}{2} = \text{constant}$ can be regarded as a wedge of angle 120° , facing a slotted polepiece composed of two planes making 240° with each other (see fig. 1 A (b)). In the gap, $\frac{\partial H^2}{\partial s}$ is constant in both longitudinal and lateral directions, while $\frac{\partial H}{\partial s}$ is constant in the longitudinal and varying in the lateral direction.

(3) *Electric and Magnetic Fields with $\frac{\partial H}{\partial s}$ constant.*

In passing through a field with $\frac{\partial H}{\partial s}$ constant in the lateral direction, a molecule in a beam would undergo a lateral deflexion,

$$s = \frac{\mu_e}{2mv^2} \cdot \frac{\partial H}{\partial s} \cdot l_1^2 \left(1 + \frac{2l_2}{l_1} \right)$$

in the usual notation, if the molecule had a fixed moment μ_e . Such a field may be obtained by using an electromagnet with cylindrical polepieces, whose curves of normal section have the polar equation $r^2 \cos 2\theta = \text{constant}$. To prove this, assume that the faces of the polepieces are equipotentials. Employ the conformal transformation

$$W = U + iV = Cz^2 = Cr^2 e^{i2\theta},$$

where r, θ are the polar co-ordinates of any point in figs. 1 B (a) or B (b), i. e.

$$U + iV = Cr^2 \cos 2\theta + iCr^2 \sin 2\theta.$$

Here take U as the potential, V the conjugate function, r, θ , the polar coordinates in the z -plane. Hence, in this case,

$$U = Cr^2 \cos 2\theta \text{ is the potential.}$$

The equipotential curves are a set of rectangular hyperbolas with the polar equation $r^2 \cos 2\theta = A$, a constant. The radial field at any point

$$P = H_r = -\frac{\partial U}{\partial r} = 2Cr \cos 2\theta.$$

The component field perpendicular to the radius vector at P is

$$H_\theta = -\frac{1}{r} \frac{\partial U}{\partial \theta} = -2C \sin 2\theta.$$

The resultant field squared $= H^2 = H_r^2 + H_\theta^2 = 4C^2 r^2$.

Hence $H = \pm 2Cr$ and $\frac{\partial H}{\partial r} = \pm 2C$. Also $\frac{\partial H}{r \partial \theta} = 0$.

The field varies only along the radius vector and its gradient $\pm 2C$ is constant. We may write

$$\frac{\partial H}{\partial s} = \left\{ \left(\frac{\partial H}{\partial r} \right)^2 + \left(\frac{\partial H}{r \partial \theta} \right)^2 \right\}^{\frac{1}{2}} = \frac{\partial H}{\partial r} = \pm 2C.$$

The lines of force, a set of orthogonal rectangular hyperbolas, are given by the equation $V = Cr^2 \sin 2\theta = \text{constant}$. Other interesting relations

are $\frac{\partial H^2}{\partial r} = 8C^2 r$, and $\frac{\partial H^2}{r \partial \theta} = 0$. The above well-known curves are depicted

in Jeans' 'Electricity and Magnetism,' 4th edition, 1920, page 266. The equipotentials $r^2 \cos 2\theta = A$ are to be regarded as the normal sections of infinitely deep cylindrical polepieces or condenser plates, according as we consider the magnetic problem or the corresponding electrical one.

The force on the molecule is proportional to $\frac{\partial H}{\partial s}$ and is directed along a radius vector, towards or away from the origin, according to the signs of the poles or charges.

Three types of polepieces or condenser plates can be constructed, all obeying the equation $U = Cr^2 \cos 2\theta = \text{constant}$. These are represented in fig. 1 B (a), B (b), and B (c). Thus in fig. 1 B (a) the system includes four polepieces or condenser plates. One of these, in the limit, may be a wedge of angle 90° . Neighbouring poles or plates are unlike in sign. In fig. B (b) is represented the second case, which consists of two polepieces or plates. One of these is a "wedge" of angle 270° . If we require a field with H large and $\frac{\partial H}{\partial s}$ constant in a lateral direction, we

may select two of the right-hand curves of fig. 1 B (b), as templates for the construction of our polepieces or condenser plates. If we require

H small and $\frac{\partial H}{\partial s}$ constant we may select two curves closer to the wedge

of fig. 1 B (b), for at the point O, $H = 0$ and since $H = 2Cr$ the field grows as we recede from O. Fig. 1 B (c) represents a third possible system of polepieces or condenser plates with a gap in which there is no lateral

variation of $\frac{\partial H}{\partial s}$, also derived from $U = Cr^2 \cos 2\theta = \text{constant}$. Thus we

may have a plane polepiece facing a slotted plane polepiece or one slotted plane facing another.

The plane faces represent $r^2 \cos 2\theta = A$ when $A=0$, the slotted faces when $A \neq 0$. In this field also $\frac{\partial H}{\partial s} = \text{constant}$ at every point and the mechanical force on a particle with fixed magnetic or electric moment is everywhere radial in direction. The origin O is a point of maximum H . It would seem that theory predicts that no electric charges or magnetic poles will be distributed further down inside the slots than the points where planes through O , making angles of $\pm 45^\circ$ with Ox , intersect the equipotentials. This implies no fields as well, below these intersecting planes.

In all the above cases there will still be "fringing field" or edge effects, *i. e.* regions traversed by the molecular beam just before entry to and just after exit from the gap proper, in which $\frac{\partial H^2}{\partial s}$ or $\frac{\partial H}{\partial s}$ will vary in ways not yet calculated.

Condensers and magnetic polepieces of these shapes have not yet been constructed.

(4) *A Note on the Gerlach-Stern Polepieces.*

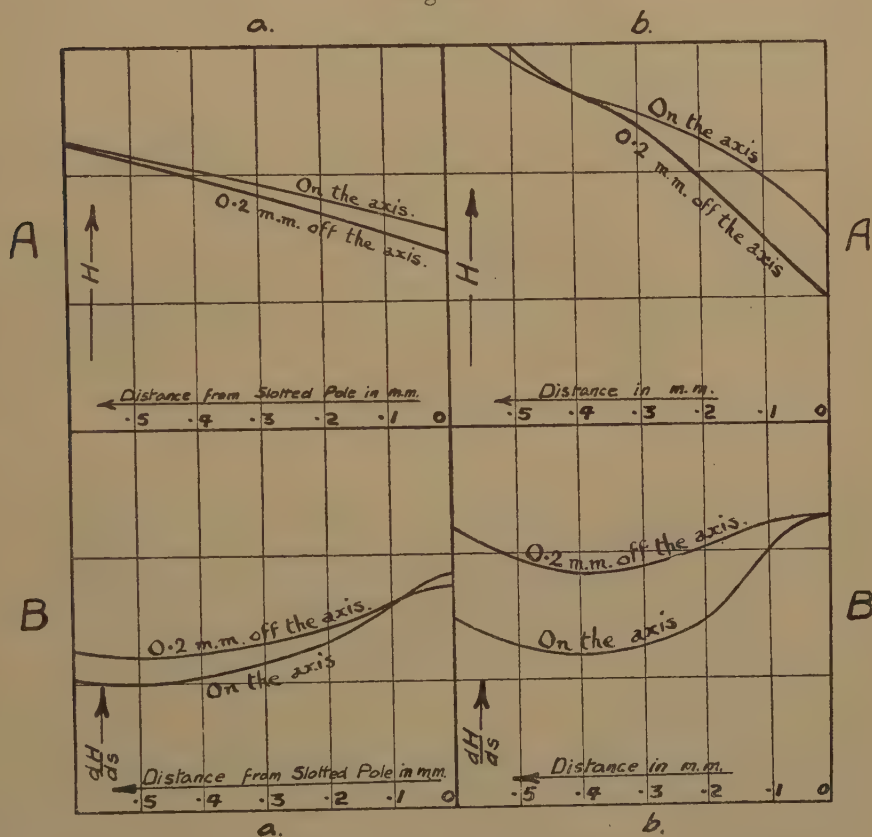
No thorough theoretical calculation of the magnetic field and its gradients in the gap of a Gerlach-Stern system of polepieces, seems to have been published. Gerlach and Stern themselves calculated by Stefan's method that part of the field due to the wedge, but did not show how to get the contribution of the slotted polepiece. In view of the want of a full theory, it is perhaps worth mentioning that the qualitative features of the changes in H and $\frac{\partial H}{\partial s}$ in such a field can be

derived by replacing the actual polepieces by line distributions. Thus, take the magnet of Estermann and Stern (1933), in which the edge of the wedge is 1 mm. away from the centre of a slot 1 mm. wide in a plane polepiece. Assume three line distributions coinciding respectively with the three edges of the system, *viz.*, a north line distribution of strength $2m$ e.m.u. coinciding with the edge of the wedge, and two south line distributions, each of strength m e.m.u. coinciding with the two edges of the slot respectively. This seems reasonable considering the symmetry of the system and the very narrow gap concerned, for the large surface densities and total pole strengths lying along the edges appear to be the dominating factors in deciding the field strengths in their vicinity. First calculate the field at any point along the axis of symmetry Os of a normal section (fig. 1 A (a)). This is, of course, in the central plane of symmetry. As we proceed from the wedge to the middle of the slot the field decreases though not linearly (see fig. 2 A (b)). Its gradient first decreases, then reaches a minimum about 0.6 mm. from the wedge, then rises to a maximum near the plane of the opening

of the slot (see fig. 2 B (b)). Next calculate the field at any point in a straight line parallel to the axis in a plane parallel to, but 0.2 mm. from, the plane of symmetry. Near the wedge the field is greater than that in the plane of symmetry but decreases more quickly (see fig. 2 A (b)). The H, s graph crosses that of the plane of symmetry when $s=0.4$ mm.

approx. The graph of $\frac{dH}{ds}$, s in the second plane shows the same feature of a decrease of $\frac{dH}{ds}$ to a minimum at about $s=0.6$ mm. followed by a

Fig. 2.



Aa, Ba. Experimental values, Gerlach & Stern type of polepieces.
Ab, Bb. Calculated values, simple model.

The distance from the slotted pole $= 1-s$ millimetres where s = distance from the tip of the wedge.

rise (see fig. 2 B (b)). The two gradient graphs intersect when $s=1.0$ mm. approx., i.e. near the mouth of the slot. The experimental graphs of

H, s , and $\frac{dH}{ds}$, s , published by Estermann and Stern, display all these features qualitatively, but (i) the separation of the two H, s graphs is

much less than with the present simple model, *e. g.* where $s=1.0$ mm., and (ii) the two gradient curves $\frac{dH}{ds}$ s , intersect when $s=0.9$ mm. as against $s=1.0$ mm. with the simple model. This can be seen by comparing fig. 2 A (a) with fig. 2 A (b), and fig. 2 B (a) with fig. 2 B (b).

I have great pleasure in thanking my colleague, Mr. J. G. Rolinson, for drawing the figures both for the present and for the earlier paper.

References.

- (1) E. McMillan, 'Physical Review,' vol. xxxviii. p. 1568 (1931); and vol. xlii. p. 905 (1931).
- (2) N. Davy, Phil. Mag. vol. xxxiii. p. 575 (1942).

24th February, 1943.

XCI. Viscosities and Rheochors of Nitric Acid, Nitroparaffins and their Isomeric Nitrites.

By J. NEWTON FRIEND and WILLIAM D. HARGREAVES*.

[Received September 2, 1943.]

THE introduction of a co-ordinate bond into a molecule produces a larger external electrical field than would be caused by ordinary co-valent bonds. Hence, such compounds are less volatile than their isomerides containing only co-valent links. This was illustrated by Sidgwick⁽¹⁾, who compared the boiling points of nitroparaffins, $R-N\begin{smallmatrix} \nearrow O \\ \searrow O \end{smallmatrix}$ and their isomeric nitrites, $R-O-N=O$, showing that the former are appreciably higher than the latter.

For a similar reason one would expect the densities, D , viscosities and latent heats at comparable temperatures to be greater for the nitroparaffins than for the corresponding nitrites. That such is the case is shown by the results of the present research, as summarized in Table I., the temperatures of comparison being the boiling points.

The viscosities are expressed as millipoises, $10^3\eta$; in the absence of other data the approximate latent heats of evaporation per gram, l , have been evaluated from Friend's equation⁽²⁾:

$$l = 2.303 \times \frac{4R}{M} \cdot \frac{T_1 T_2}{T_2 - T_1} \log(\eta_1/\eta_2) \\ = 2.303 \times 4RB/M,$$

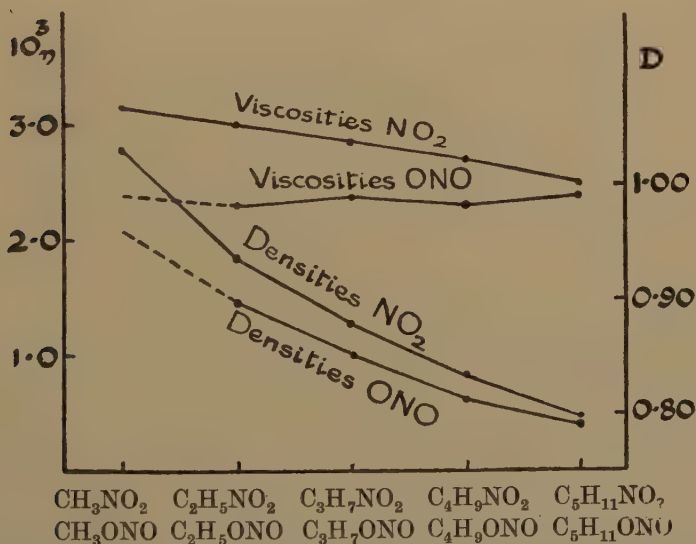
* Communicated by the Authors,

where M is the molecular weight and R the gas constant, taken as 1.985, l being given in gram calories. The constant B is taken from Sheppard's equation ⁽³⁾, $\log 10^3 \eta = A + B/T$.

TABLE I.—Boiling Point Data.

	B. pt. °C.	D.	$10^3 \eta$.	l .
HONO ₂	80	1.4128	4.55	160*
CH ₃ NO ₂	100	1.0272	3.13	132†
C ₂ H ₅ NO ₂	115	0.9317	3.01	116
C ₂ H ₅ ONO	18	8961	2.28	75
<i>n</i> C ₃ H ₇ NO ₂	131	0.8757	2.87	104
<i>n</i> C ₃ H ₇ ONO	50	8503	2.38	77
<i>n</i> C ₄ H ₉ NO ₂	152	0.8305	2.70	93
<i>n</i> C ₄ H ₉ ONO	76.5	8105	2.31	74
<i>n</i> C ₅ H ₁₁ NO ₂	172.5	0.7962	2.50	88
<i>n</i> C ₅ H ₁₁ ONO	104.5	7902	2.39	69
<i>iso</i> C ₅ H ₁₁ NO ₂	163.5	0.8104	2.88	94
<i>iso</i> C ₅ H ₁₁ ONO	97	7865	2.35	65

Fig. 1.



* Int. Critical Tables, v. p. 136, give 115 as the value for acid boiling at 86° C., but state that this is certainly too low.

† Mathews (J. Amer. Chem. Soc. xlviii. p. 562 (1926)) found 134.94 experimentally.

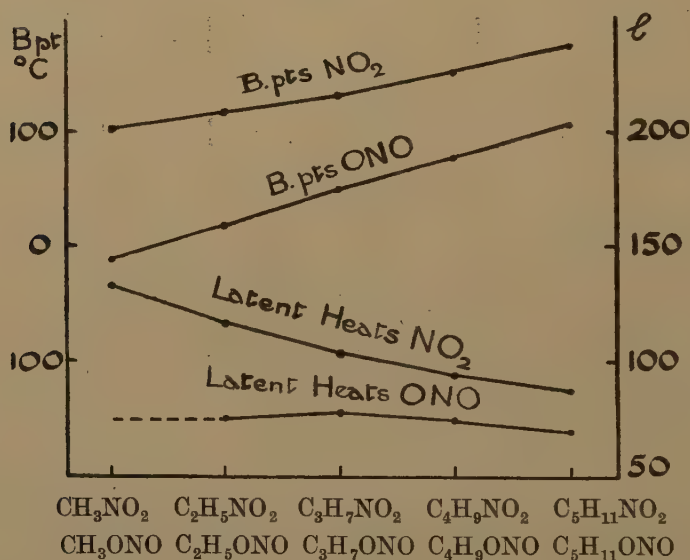
These results are shown graphically in figs. 1 and 2. As the molecular weight increases the differences between the isomerides become less pronounced. Whilst the viscosities and latent heats of the nitro-paraffins steadily fall, those of the nitrites are fairly constant. As usual, the boiling points of the *iso*-derivatives are lower than those of the normal.

It has already been shown ⁽⁴⁾ that the rheochor, *R*, of organic substances is both additive and constitutive. The rheochor is defined as

$$M(10^3\eta)^{\frac{1}{2}}/(D+2d),$$

where *d* is the density of the vapour. Mean values of *R* already determined are C=12.8, O=10 and H=5.5. The data in Table I. enable values to be assigned to the NO₂ and ONO groups respectively, the

Fig. 2.



effect of the co-ordinate bond being the difference between the two. The results are given in Table II.

The mean rheochors are thus: NO₂=38.9; ONO=39.3, the co-ordinate bond inducing a fall of 0.4 unit. The rheochor of free N₂O₄, using the data of Thorpe and Rodger ⁽⁵⁾, is 2×37.9.

Concentrated nitric acid is usually regarded as consisting essentially of undissociated HONO₂. The presence of the NO₂ group is indicated by the Raman effect, etc. The rheochor should thus be

$$5.5+10+38.9=54.4.$$

The values found are :

Temp. °C.	18.2	35.85	55.2	69.0	(80)
R(HNO ₃)	54.6	54.1	53.8	53.8	(53.7)

In view of the tendency of the concentrated acid to decompose as the temperature rises the agreement appears satisfactory. The figures in parentheses refer to the boiling point and have been obtained by extrapolation.

TABLE II.—Rheochors at the Boiling Point.

	R (Compound).	R (NO ₂ or ONO).	R (co-ord. bond).
CH ₃ NO ₂	68.3	39.0	—
C ₂ H ₅ NO ₂	92.0	38.9	—0.3
C ₂ H ₅ ONO	92.3	39.2	
<i>n</i> C ₃ H ₇ NO ₂	115.4	38.5	—0.5
<i>n</i> C ₃ H ₇ ONO	115.9	39.0	
<i>n</i> C ₄ H ₉ NO ₂	139.6	38.9	—0.5
<i>n</i> C ₄ H ₉ ONO	140.1	39.4	
<i>n</i> C ₅ H ₁₁ NO ₂	163.7	39.2	—0.1
<i>n</i> C ₅ H ₁₁ ONO	163.8	39.3	
<i>iso</i> C ₅ H ₁₁ NO ₂	163.6	39.1	—0.5
<i>iso</i> C ₅ H ₁₁ ONO	164.1	39.6	

Experimental.

Fuming nitric acid and nitromethane were obtained from the British Drug Houses. The other nitro-paraffins were prepared from the alkyl bromides ⁽⁶⁾, which had been synthesized from the corresponding redistilled alcohols ⁽⁷⁾.

The alkyl nitrites were prepared from the corresponding alcohols by the action of sodium nitrite and sulphuric acid, the temperature being kept below 2° C. during addition of acid. It was observed that, when the concentrated acid and undiluted alcohol were used, the product distilled over a considerable temperature range, the boiling point of the last 5 to 10 per cent. corresponding to that of the isomeric nitroparaffin. When the alcohol was diluted prior to the addition of acid, the product distilled almost entirely at the boiling point of the nitrite.

Published data for the densities and boiling points of some alkyl nitrites suggest that the liquids were contaminated with some of the denser, high-boiling fraction. Thus, the boiling point of *n*-propyl nitrite, which we found to be 50° C. at 753 mm., is given by Cahours ⁽⁸⁾ as 57°. This author gives the density at 21° as 0.935 (0.8832), whilst Lowenherz ⁽⁹⁾ gives 0.8844 (0.8843) at 20°, the figures in parentheses being our results. Pribram ⁽¹⁰⁾ gives 0.9981 as the density at 0° C., our extrapolated value being 0.9070.

Nitric acid.—The fuming acid was distilled three times from solution in concentrated sulphuric acid in an all-glass apparatus, filtered through a sintered glass funnel and aerated by passage of dry air until pale yellow. As the temperature rose during the course of the experiments the colour changed to red, and above 70° C. brown fumes were evolved so that viscosity measurements were discontinued. The liquid boiled at 80 to 85° C. under 753.5 mm. $A = -0.9004$; $B = 550.0$. $l = 160$.

Temp. °C.	18.2	35.85	55.2	69.0	(80)
D	1.5335	1.4998	1.4629	1.4351	(1.4128)
$10^3\eta$	9.729	7.538	5.950	5.109	(4.55)

R (at the boiling point) = 53.7; calc. 54.4.

Nitromethane.—Dried over calcium chloride and twice distilled, the middle fraction of the second distillate being collected direct in the viscometer and density bottle. B. pt. 99.0° to 100.2° C. at 743.8 mm. Density a linear function of the temperature, $D = 1.1644 - 1.37 \times 10^{-3}t$. The data agree closely with those of Boyd and Copeland⁽¹¹⁾ over their range (25° to 60° C.), but differ widely from those of Philip and Oakley⁽¹²⁾, although the viscosity data between 40° and 80° show good agreement with those of the latter authors, and with the equation given in the International Critical Tables⁽¹³⁾. $A = 0.6806$; $B = 439.2$. $l = 132$.

Temp. °C. .	16.4	42.7	61.8	78.6	96.3	(100)
D	1.1416	1.1063	1.0798	1.0563	1.0318	(1.0272)
$10^3\eta$	6.920	5.134	4.280	3.699	3.219	(3.13)

$R = 68.3$; calc. 68.2.

Nitroethane.—B. pt. 113.8° to 115° C. at 749.8 mm. Density almost linear with temperature, data agreeing closely with those of Boyd and Copeland. $A = -0.7526$; $B = 478.0$. $l = 116$.

Temp. °C. .	18.6	38.5	56.8	76.9	92.9	108.5	(115.0)
D	1.0482	1.0242	1.0018	0.9776	0.9583	0.9395	(0.9317)
$10^3\eta$	7.704	6.040	4.966	4.095	3.579	3.162	(3.01)

$R = 92.0$; calc. 92.0.

Nitropropane.—B. pt. 130.0° to 131.0° C. at 755.5 mm. Density almost linear with temperature, data agreeing closely with those of Boyd and Copeland. $A = -0.7940$; $B = 505.8$. $l = 104$.

Temp. °C. .	15.3	53.4	76.9	92.7	107.2	122.0	(131.0)
D	1.0053	0.9648	0.9386	0.9204	0.9034	0.8861	(0.8757)
$10^3\eta$	9.253	5.742	4.513	3.916	3.450	3.066	(2.87)

$R = 115.4$; calc. 115.8.

Nitro n-butane.—B. pt. 150° to 152° C. at 756.4 mm. Densities slightly lower than those of Boyd and Copeland. $A = -0.8082$; $B = 526.5$. $l = 93$.

Temp. °C. .	15.2	51.1	74.2	96.45	113.5	142.7	(152.0)
D	0.9734	0.9385	0.9155	0.8932	0.8749	0.8412	(0.8305)
$10^3\eta$	10.55	6.56	5.11	4.145	3.592	2.872	(2.70)

$R = 139.6$; calc. 139.6.

Nitro-n-pentane.—B. pt. 172.5° C. at 758.5 mm. $A = -0.8708$;
 $B = 564.9$. $l = 88$.

Temp. °C. .	21.45	60.1	80.8	100.5	121.1	140.7	161.0	(172.5)
D	0.9472	0.9087	0.8882	0.8686	0.8480	0.8282	0.8078	(0.7962)
$10^3\eta$	11.31	6.68	5.31	4.355	3.65	3.112	2.696	(2.50)

$R = 163.7$; calc. 163.4.

Nitro iso-pentane or 3 methyl nitro *n*-butane.—B. pt. 163.5° C. at 758.5 mm. $A = -0.9231$; $B = 603.2$. $l = 94$.

Temp. °C. .	17.4	40.2	60.25	81.2	100.1	122.0
D	0.9600	0.9378	0.9171	0.8957	0.8762	0.8537
$10^3\eta$	19.45	10.07	7.70	6.012	4.926	4.018
Temp. °C.			140.7	155.0	(163.5)	
D			0.8342	0.8193	0.8104	
$10^3\eta$			3.422	3.062	(2.88)	

$R = 163.6$; calc. 163.4.

Ethyl nitrite.—B. pt. 18.1° C. at 737.7 mm. $A = -0.7047$; $B = 309.7$.
 $l = 75$.

Temp. °C.	0.2	9.97	14.3	(18.1)
D	0.9223	0.9081	0.9017	(0.8961)
$10^3\eta$	2.683	2.454	2.365	(2.28)

$R = 92.3$; calc. 92.4.

n-Propyl nitrite.—B. pt. 50.0° C. at 753.1 mm. $A = -0.7803$; $B = 373.5$. $l = 77$.

Temp. °C.	16.5	29.45	38.25	45.9	(50.0)
D	0.8883	0.8737	0.8637	0.8550	(0.8503)
$10^3\eta$	3.234	2.848	2.632	2.461	(2.38)

$R = 115.9$; calc. 116.2.

n-Butyl nitrite.—B. pt. 75.0° to 76.5° C. at 737.9 mm. $A = -0.8325$;
 $B = 418.2$. $l = 74$.

Temp. °C. .	16.4	46.0	54.7	60.9	71.0	(76.5)
D	0.8862	0.8489	0.8383	0.8302	0.8172	(0.8105)
$10^3\eta$	4.096	2.990	2.774	2.626	2.407	(2.31)

$R = 140.1$; calc. 140.0.

n-Amyl nitrite.—B. pt. 104.0° to 104.5° C. at 759.6 mm. $A = -0.7940$;
 $B = 442.4$. $l = 69$.

Temp. °C. ...	18.6	39.2	58.9	75.1	90.7	99.1	(104.5)
D	0.8809	0.8592	0.8386	0.8212	0.8050	0.7961	(0.7902)
$10^3\eta$	5.321	4.207	3.461	3.000	2.647	2.486	(2.39)

$R = 163.8$; calc. 163.8.

Our density data agree well with those given by Sugden ⁽¹⁴⁾.

iso-Amyl nitrite.—B. pt. 95.5° to 97.0° C. at 739.7 mm. $A = -0.7585$;
 $B = 418.1$. $l = 65$.

Temp. °C.	12.7	42.7	52.7	64.6	78.5	(97.0)
D	0.8829	0.8481	0.8363	0.8226	0.8072	(0.7865)
$10^3\eta$	5.066	3.684	3.351	3.014	2.693	(2.35)

$R = 164.1$; calc. 163.8.

Our density data agree very satisfactorily with those found by Sugden ⁽¹⁴⁾ and his co-workers.

References.

- (1) Sidgwick, 'The Electronic Theory of Valency' (1927).
- (2) Friend, Trans. Faraday Soc. xxxi. p. 542 (1935).
- (3) Sheppard, J. Rheology, i. p. 208 (1930).
- (4) Friend, 'Nature,' cl. p. 432 (1942); Friend and Hargreaves, Phil. Mag. xxxiv. p. 643 (1943).
- (5) Thorpe and Rodger, Phil. Trans. clxxxv. p. 45 (1895).
- (6) Reynolds and Adkins, J. Amer. Chem. Soc. li. p. 279 (1929).
- (7) Kamm and Marvel, *ibid.* xlii. p. 299 (1920).
- (8) Cahours, *Compt. Rend.* lxxvii. p. 745 (1873).
- (9) Lowenherz, *Zeitsch. physikal. Chem.* vi. p. 552 (1890).
- (10) Pribram and Handl, *Monatsh.* ii. p. 643 (1881).
- (11) Boyd and Copeland, J. Amer. Chem. Soc. lxiv. p. 2540 (1942).
- (12) Philip and Oakley, J. Chem. Soc. cxxv. p. 1189 (1924).
- (13) International Critical Tables, iii. p. 28.
- (14) Sugden, Reed and Wilkins, J. Chem. Soc. cxxvii. p. 1525 (1925).

The Technical College,
Birmingham.

XCII. *Unreleased Energy in Hydrocarbon Flame Gases.*

By W. T. DAVID, Sc.D., and J. MANN, B.Sc.*

[Received September 20, 1943.]

It has been shown by means of experiments made during the pre-pressure period in large closed vessel explosions that the gases resulting from the combustion of carbon monoxide-air mixtures contain a large amount of latent energy, and in virtue of this are—except in the case of very over-rich mixtures—abnormally dissociated †. It has also been shown that flame gases resulting from the combustion during the pre-pressure period of moist carbon monoxide-air mixtures attain higher temperatures than those from dry mixtures, and therefore contain less of this unreleased energy ‡, that is to say, energy unreleased in the flame gases for the purpose of increasing their temperature. The experiments described in this paper were carried out with the hydrocarbon gases which we had available in order to determine the extent to which latent energy and abnormal dissociation obtain in open flames. Moist and dry mixtures of these with air were burnt at atmospheric pressure in a specially

* Communicated by the Authors.

† David, Leah, and Pugh, Phil. Mag. xxxi. p. 156 (1941); David, Proc. I. Mech. E. (in press).

‡ David and Pugh, 'Nature' cl. p. 1098 (1939); cl. p. 521 (1942); cli. p. 392 (1943).

constructed burner and the flame temperatures were measured by resistance thermometry, using very fine quartz-coated platinum-rhodium wires of 0.0005 inches overall diameter.

The gases used were methane, ethane, propane, butane, ethylene and butylene of varying degrees of purity, and analyses of these are given in the Appendix.

Description of Apparatus.

The gaseous mixtures were made up prior to combustion and stored in a tank: the combustible gases were taken direct from the cylinder and the air was dried by passing through a tube containing phosphorus pentoxide. Thorough mixing of the gases was ensured by means of a fan mounted inside the tank, and also by standing. The mixture from the tank was led to the burner through a flame trap and a small drying tube. (The drying tube was used principally as a check on the one used when making up the mixtures). When moist mixtures were required, the gases from the tank were passed through a flask containing water and packed with cotton wool. Since the mixture passed through the water in the form of very small bubbles and then went through the moist cotton wool, it was assumed for the purpose of calculating the mixture composition that the gases were saturated with water vapour. The burner was of special design and consisted of a copper tube 1 inch in diameter flattened for a length of 2 inches, the internal dimensions then being $1\frac{1}{2}$ in. \times $\frac{3}{32}$ in. A cap, in which the orifice was cut, was soldered to the end of this tube. The most satisfactory orifice was found to be a parallel slit $\frac{1}{16}$ in. wide and about $1\frac{5}{16}$ in. long, at the ends of which were two $\frac{1}{16}$ in. square holes. These two smaller holes "tied down" the ends of the flame, and gave it stability when very weak mixtures were being burnt. The burner was enclosed in a rectangular open topped chamber with mica windows. No cooling of the burner was attempted as none was found to be necessary since the burner was lit only for short periods. Suspended above the burner was the thermometer, which consisted of a quartz-coated platinum-rhodium (10 per cent. rhodium) wire about one inch in length and 0.0005 inches in diameter, mounted on two stout platinum-rhodium leads which, together with compensating leads, were secured in a holder. The thermometer wire was capable of easy and rapid adjustment in position both horizontally and vertically, so that it could be placed in the required zone of the flame. In the experiments reported in this paper, this was just above the inner cone. The resistance of the wire was measured by means of a balanced bridge arrangement and the temperatures were calculated as for previous closed vessel experiments, using a furnace calibration.

Experimental Results.

The temperatures attained by the quartz-coated wires and the calculated unreleased energies for methane and ethane are shown in fig. 1, for propane and butane in fig. 2, and for ethylene and butylene

in fig. 3. The wire temperatures shown have not been corrected for radiation loss, but in calculating the unreleased energies a generous correction has been applied. The percentage of combustible gas in the

Fig. 1.

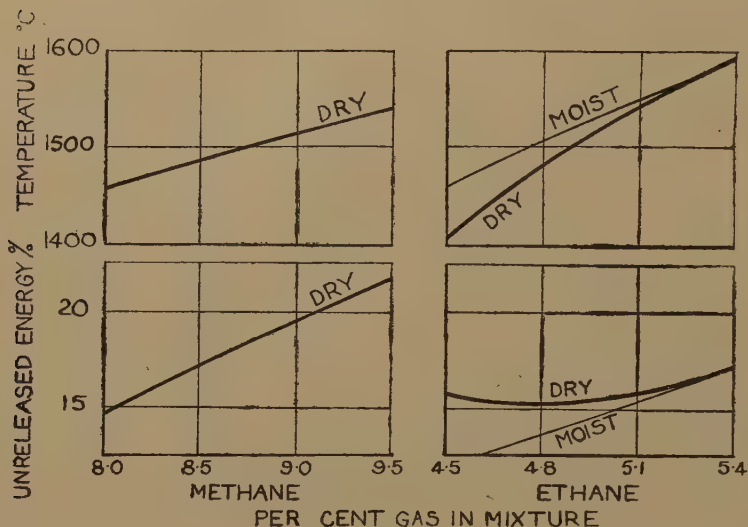
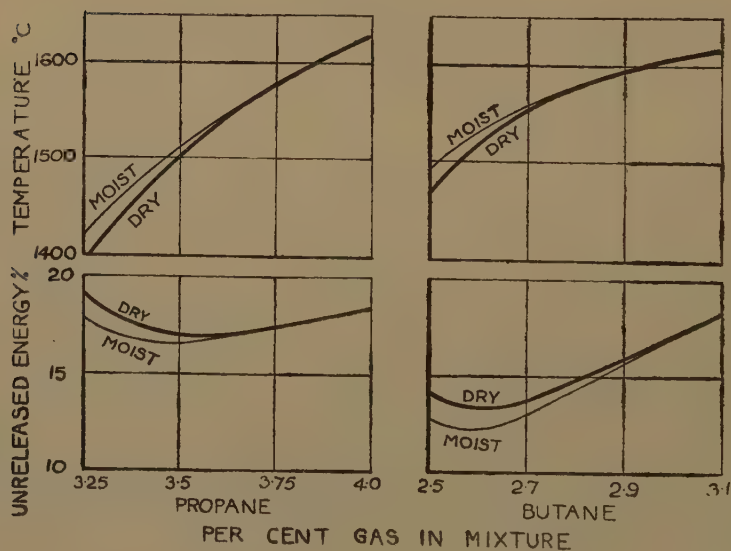


Fig. 2.



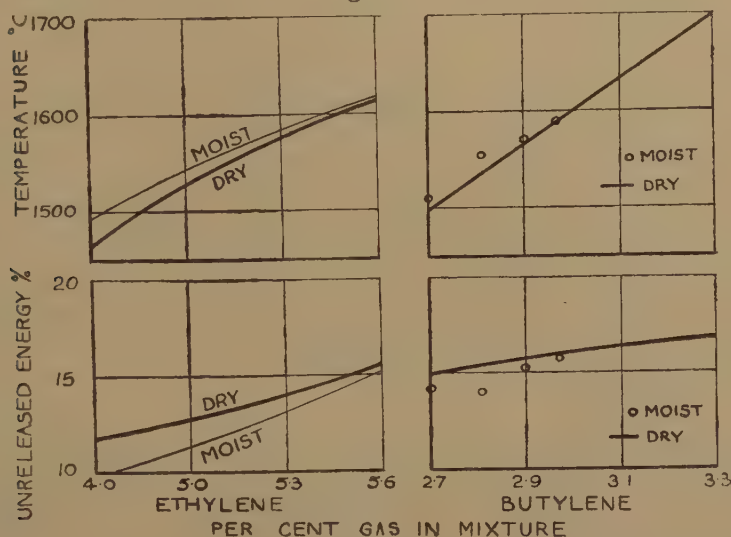
mixture, as shown on the abscissæ scales, is the total percentage of gas as taken from the cylinder, except in the case of methane where the mixture strength has been corrected for impurities.

Discussion.

Flame gases contain a long-lived latent energy, which probably resides in a proportion of the newly-formed triatomic molecules in an abnormal structural condition. As a result of this abnormal dissociation obtains in the flame gases*. The temperature attained by a wire immersed in flame gases is therefore the net result of heating due to :—

- (1) the temperature of the flame gases, which may be assumed to be proportional to their mean molecular translational energy ;
- (2) the unloading of the latent energy, or the normalizing of the molecules of abnormal structure on its surface ; and

Fig. 3.



- (3) the recombination upon its surface of the abnormally dissociated molecules—provided the surface is of a type which promotes surface combustion,

and of cooling due to :—

- (4) radiation from the wire ; and
- (5) conduction into the cooler supporting leads.

In the case of the quartz covered wires which we have used throughout the experiments, there was no recombination of the abnormally dissociated molecules and therefore no heating due to (3)†.

* David, Leah and Pugh, *Phil. Mag.* xxxi. p. 156 (1941) ; and David, *Proc. I. Mech. E.* (in press).

† It was hoped that experiments with very fine plain platinum rhodium wires could be carried out, but it was found that the plain wires fused or broke down even in weak flames. For example, the flame gases from a mixture containing 3.6 per cent. of propane fused a plain wire when the coated wire temperature was only 1540° C.

The thin wires (of overall diameter 0.0005 inches) were about one inch long, and the effects of (5) were therefore negligible.

The wire may thus be said to give a measure of (1) subject to a correction for a cooling due to radiation, less a heating resulting from the unloading of latent energy. It is not possible to frame an estimate of the amount of heating due to the unloading of latent energy. It is possibly relatively small and will be disregarded*.

The loss due to radiation is difficult to estimate, but it is based upon the assumption that when the wire temperature is 1400° C. a correction of 50° C. should be applied. There is much in our work to suggest that this is on the high side, and when it is recalled that the effects of heating due to the unloading of latent energy have been disregarded, it may be assumed that the wire temperatures, when corrected for radiation loss, yield in full measure a temperature which may be regarded as being proportional to the mean molecular translational energy in the flame gases. On the basis of this temperature, the specific heats of the gases constituting the flame gases and the heat of combustion, that proportion of the heat of combustion which is unreleased in the flame gases (and exists as latent energy *plus* abnormal dissociation) may be calculated. These calculations were made allowing for the impurities in the gases but without allowance for normal dissociation†.

It will be noticed that the unreleased energies vary from about 10 per cent. to more than 20 per cent. of the heat of combustion. It is least in the flame gases from weak moist ethylene mixtures and greatest in those from the theoretically correct methane mixture. It will be also noticed that in the case of several of the gases the unreleased energy at first decreases, and we think it highly probable that had we been able to use weaker mixtures, this would have been the case for all the gases. Later, the unreleased energy steadily increases and is greatest when the

* In earlier work by one of us (David, *Phil. Mag.* xxiii. p. 251 (1937)), with thin uncoated platinum wires, it was thought that the heating due to the unloading of latent energy balanced—indeed more than balanced—the radiation loss. This conclusion was arrived at as a result of a comparison of the temperatures attained by the wires with those measured by the sodium line reversal method. There is no doubt that the uncoated platinum wire temperatures were far too high (even though no radiation correction was applied), but it is now realized that this was mainly due to the recombination on the wire surface of the abnormally dissociated molecules.

† The flame gases lose a small amount of heat by radiation during the short interval while they pass from the combustion zone to the point of temperature measurement. This amounts only to about 1 per cent., or at most 2 per cent. of the heat of combustion, and has been allowed for in making the calculations. Normal dissociation can have little effect upon any of the results as the maximum temperature recorded is in the region of 1650° C. In the case of ethane, for example, the amount of unreleased energy allowing for normal dissociation was found to vary from a fraction of a per cent. for a mixture containing 4.5 per cent. of ethane to a little more than one per cent. for the combining proportions mixture.

combustible gas-air mixture strengths are theoretically correct. The explanation would appear to be as follows.

It is known that as the natural rate of burning of the inflammable mixture is increased, the instantaneous pressure and temperature in the flame front are increased * and the latent energy consequently decreased. The latent energy part of the unreleased energy in the open flame gases therefore decreases with mixture strength. The other part of the unreleased energy, viz. that due to abnormal dissociation, decreases as the latent energy decreases, but on the other hand increases as the flame gas temperature increases—slowly at the lower temperatures and more rapidly at the higher temperatures. It would appear, therefore, that in the case of the weaker mixtures the decrease in the unreleased energy curves is due to the decrease in latent energy and in the case of the stronger mixtures the increase is due to increasing abnormal dissociation.

The effects of introducing water vapour into the mixtures were observed. It was found that for the majority of the gases used, the moist mixtures yielded temperatures higher than the dry ones in the case of weak mixtures, but as the mixture strength was increased the difference between moist and dry temperatures became less and less until mixture strengths were reached in which the proportion of the combustible gas was about 90 per cent. of the theoretical proportion, when the moist and dry temperatures were sensibly the same. This may be seen from an examination of the figures relating to ethane, propane, butane and ethylene. The influence of water vapour on the combustion of methane mixtures was not fully explored, but we believe it to be similar to that in the case of the other mixtures. Its influence on the combustion of butylene mixtures was exceptional in that it appeared to be very variable, as may be seen from the individual observations shown in the figure relating to these mixtures.

APPENDIX.

Analyses of Gases Used.

<i>Methane</i>	Methane	90.50
	Ethane	2.25
	Carbon monoxide	0.51
	Carbon dioxide	0.17
	Oxygen	0.05
	Nitrogen	6.52
		<hr/> 100.00 <hr/>

* The unreleased energies are much less in open flames than during the pre-pressure period in closed vessel explosions—except in the case of very weak mixtures. The reason would appear to be that during the pre-pressure period the propagation of flame is wholly by chain reaction mechanism and high instantaneous pressures, and temperatures in the flame front are not built up until after some distance of flame travel (Proc. I. Mech. E. (in press)).

<i>Ethane</i>	Ethane	90.0
	Methane	>2.0
	Propane }	>2.0
	Propylene }	
	Ethylene	>10.0
<i>Propane</i>	Propane	100.0
<i>Butane</i>	Butane	99+
impurity	Propane	
<i>Ethylene</i>	Ethylene	97-99%
	CH ₄ +N ₂	1-3%
	(equally)	
<i>Butylene</i>	Butylene	85.8
	Butane	12.2
	Propane	1.2
	Propylene	0.5
	Heavier H.C.'s	0.3
		<hr/> 100.0 <hr/>

XCIII. *Some Properties of a Special Type of Electrical Pulse.*

By F. F. ROBERTS, B.Sc.(Eng.) and J. C. SIMMONDS, Ph.D.*

[Received July 14, 1943.]

REPEATED pulses of current and voltage are being used to an increasing extent in varied types of electrical apparatus, and papers have appeared which give details of the Fourier analyses of a number of recurrent pulse wave-forms ⁽¹⁾ †. Other writers have shown how the Fourier integral may be employed to determine the frequency spectrum corresponding to most practical pulse shapes independently of the pulse repetition frequency ⁽²⁾. A very interesting type of pulse shape, however, does not appear to have been considered before. This pulse is described below and some of its properties are given.

The Exponential Recurrent Pulse.

It is well known that recurrent pulses can be generated in the anode current of a valve by applying a large alternating voltage to the control grid together with a sufficiently large negative bias, so that practically zero anode current flows in the absence of alternating voltage. The repetition frequency of the pulses is obviously fixed by the frequency of the alternating signal applied to the grid, whilst the duration of each

* Communicated by the Authors.

† For numbered references see Bibliography.

pulse depends in general upon the magnitude and wave-form of the alternating signal, on the bias voltage and on the valve characteristics. Many valves, and in particular certain high-slope pentodes, have a grid volts-anode current characteristic which is very nearly a straight line, when plotted with a logarithmic scale of current, over a very wide range of values. This is in accordance with the theory of valves ⁽³⁾. It will be evident from the following paragraphs that theoretical manipulation of the pulse wave-form (corresponding to the effect of various practical circuit elements) is considerably simplified if the above valve law is assumed to be exact, and if the applied alternating voltage at the control grid is taken to be sinusoidal. In fact, useful design figures, based upon this assumption, can be obtained quite readily.

Assume that the valve characteristic can be represented by the equation :—

$$I = Ae^{Bv}, \quad \dots \dots \dots (1)$$

where

I = anode current,

e = base of natural logarithms,

v = grid voltage,

A and B = constants of the valve.

Now let $v = E \cos pt - V, \quad \dots \dots \dots (2)$

where $E \cos pt$ = alternating component of the voltage applied to the valve grid,

and V = negative bias voltage applied to the grid of the valve.

Then, from equations (1) and (2),

$$I = Ae^{-BV} \cdot e^{BE \cos pt}, \quad \dots \dots \dots (3)$$

The factor Ae^{-BV} is clearly a constant for a given valve and operating conditions, whilst the factor $e^{BE \cos pt}$ determines the wave-form and also affects the amplitude; for it is clear that the peak of the pulse occurs when $\cos pt$ is unity and the peak amplitude is then

$$\hat{I} = Ae^{B(E-V)}, \quad \dots \dots \dots (4)$$

From equations (3) and (4) it is seen that :

$$\frac{I}{\hat{I}} = e^{BE(\cos pt - 1)}, \quad \dots \dots \dots (5)$$

thus, the ratio of the instantaneous current to the peak current is dependent only upon the valve constant B and the applied alternating voltage. A number of wave-forms derived from equation (5) for various values of BE are shown in fig. 1. As the constant B for suitable valves lies between 5 and 10, it will be seen that quite narrow pulses can be produced without excessive voltages being applied to the control grid of the valve.

The Frequency Components of the Exponential Pulse.

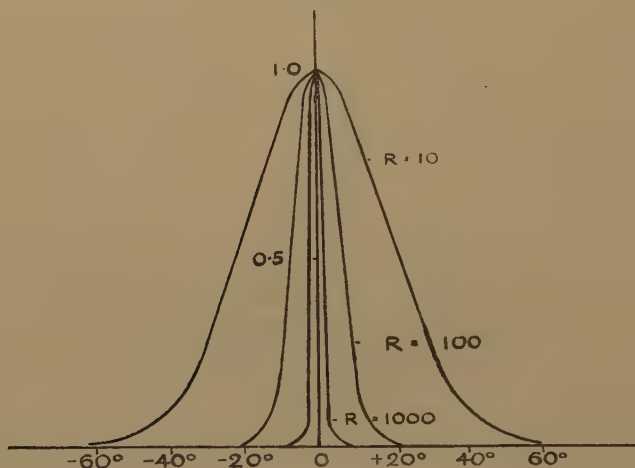
It is frequently necessary to determine the magnitudes of the frequency components of the current forming the pulse. It is obvious that these components will consist of some or all of the harmonics of the alternating voltage applied to the control grid. In the case of the exponential pulse the components are obtained immediately, for it is well known that

$$e^{BE \cos \theta} = I_0(BE) + 2 \sum_{n=1}^{\infty} I_n(BE) \cos n\theta, \quad \dots \quad (6)$$

where $I_n(BE)$ is the modified Bessel function of the first kind of order n .

Inspection of tables of these functions shows that, although for a given order n the values increase fairly rapidly with the argument BE ,

Fig. 1.

Wave-form of exponential pulse ($e^{R(\cos \theta - 1)}$).

yet for a given argument the values decrease nearly as rapidly with increasing order n . Unfortunately, tables of these functions do not appear to have been extended to orders higher than the second for arguments greater than 10—and it is just these higher order functions that are of particular interest in the present connexion. Values can, of course, be computed from the asymptotic series

$$I_n(x) = \frac{e^x}{\sqrt{2\pi x}} \left\{ 1 - \frac{(4n^2 - 1)}{1! \cdot (8x)} + \frac{(4n^2 - 1)(4n^2 - 3^2)}{2! \cdot (8x)^2} \dots \right\} \quad \dots \quad (7)$$

An obvious approximation to $I_n(x)$ is, from equation (7),

$$I_n(x) \doteq \frac{e^{\left(x - \frac{n^2}{2x}\right)}}{\sqrt{2\pi x}}, \quad \dots \quad (8)$$

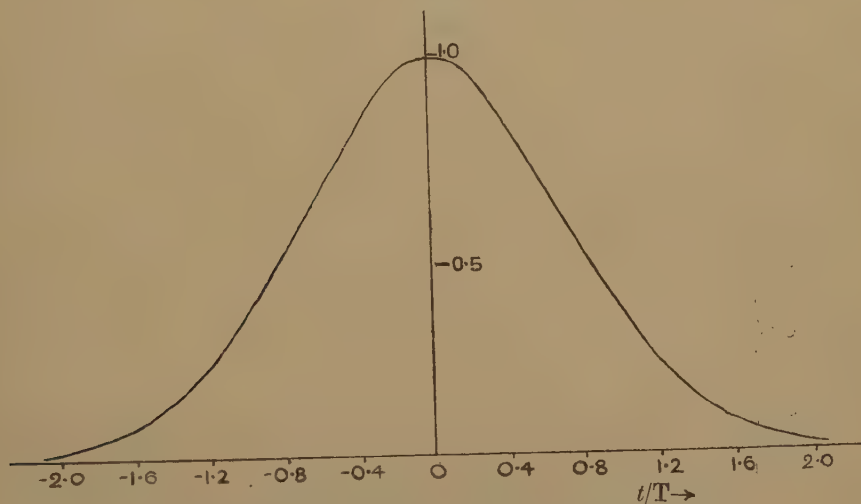
but it is not easy to determine the degree of approximation for general values of x and n ,

An alternative method of approach is to determine the frequency spectrum of a single unrepeated pulse of current. This frequency spectrum is then the envelope of the frequency components of the recurrent pulse. The pulse defined by equation (5) is essentially a recurrent pulse, but a single unrepeated pulse of almost identical shape can be obtained by expanding $\cos pt$ in the exponent of e in equation (5) and neglecting powers of pt higher than the second. When this is carried out a pulse defined by

$$\frac{I}{I_0} = e^{-\frac{1}{2}BE(pt)^2} \quad \dots \quad (9)$$

results. This pulse is essentially a single unrepeated pulse, centred at

Fig. 2.



Wave-form of probability function pulse ($e^{-(t/T)^2}$).

the origin, and having a breadth $2T$ at $\frac{1}{e}$ of the peak amplitude given by

$$e^{-\frac{1}{2}BEp^2T^2} = e^{-1} \quad \dots \quad (10)$$

or

$$T^2 = \frac{2}{BEp^2} \quad \dots \quad (11)$$

From equations (9) and (11) it is apparent that the pulse wave-form is given by

$$\frac{I}{I_0} = e^{-(t/T)^2}, \quad \dots \quad (12)$$

--this expression will be recognised as the probability function. A single pulse as defined by equation (12) is shown in fig. 2. It is interesting to note that the agreement between equations (5) and (12) is such that the respective wave-forms could only with difficulty be distinguished if drawn on the scale of fig. 2, so long as BE has a value of not less than 10.

Thus, to a close degree of approximation, the wave-form given by equation (5) may be considered in these circumstances to be made up of regularly repeated pulses of the wave-form given by equation (12).

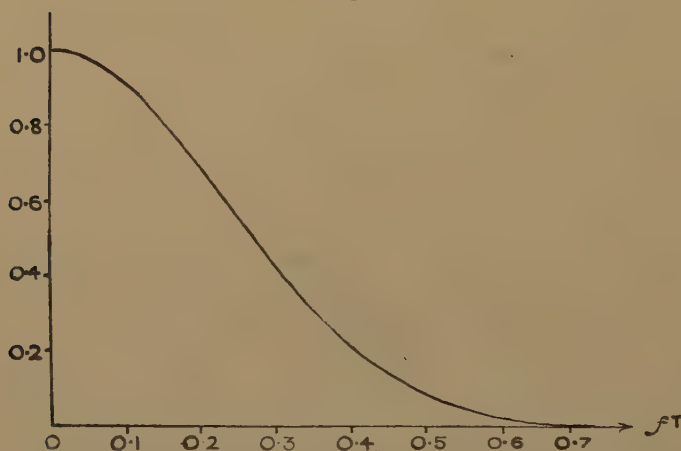
Now equation (12), considered as a function of time, fulfils the necessary conditions for the application of the Fourier transform. Therefore,

$$g(f) = \int_{-\infty}^{\infty} e^{-(t/T)^2} e^{-j\omega t} dt, \quad \dots \dots \dots (13)$$

where $g(f)$ represents the relative differential amplitude of any given frequency component f of the spectrum and ω is the angular frequency corresponding to f . Now, since the pulse given by equation (12) is obviously an even function both of time and frequency,

$$g(f) = 2 \int_0^{\infty} e^{-(t/T)^2} \cos \omega t dt. \quad \dots \dots \dots (14)$$

Fig. 3.



Frequency spectrum of probability function pulse ($e^{-\pi^2 T^2 f^2}$).

It is, therefore, easily shown that

$$g(f) = \sqrt{\pi} T e^{-\pi^2 T^2 f^2}. \quad \dots \dots \dots (15)$$

Equation (15) is illustrated in fig. 3, the constant factor $\sqrt{\pi}$ being omitted. From the theory of the Fourier transform, it follows that the amplitude of any particular harmonic of the pulse current for a given finite repetition frequency $F = p/2\pi$ is given by

$$\frac{I}{I_0} = \sqrt{\pi} F T e^{-\pi^2 T^2 F^2 n^2}. \quad \dots \dots \dots (16)$$

The expression $1/(FT)$ represents the "Relative Sharpness," *i.e.* the number of half pulse widths in the pulse repetition period. Also, from equation (11), it determines BE by the relationship

$$BE = \frac{1}{2\pi^2 (FT)^2}. \quad \dots \dots \dots (17)$$

It will be observed that equation (6) can now be written as

$$e^{\text{BE} \cos \pi t} = \frac{e^{\text{BE}}}{\sqrt{2\pi \text{BE}}} \{1 + 2 \sum e^{-n^2/2\text{BE}} \cos n\pi t\}. \quad (18)$$

It is interesting to note that this expression could have been obtained by using the approximation to $I_n(x)$ given by equation (8).

The Use of the Recurrent Exponential Pulse in Theoretical Analysis.

Some illustrations of the ease with which the exponential recurrent pulse can be manipulated mathematically will now be given.

Let $P_1 = A_1 e^{R_1 \cos(\omega_1 t + \phi_1)}, \quad (19)$

$P_2 = A_2 e^{R_2 \cos(\omega_2 t + \phi_2)}, \quad (20)$

be two given pulse wave-forms. Then the result of superimposing the two wave-forms, *e. g.* by connecting the two sources in parallel, is simply given by $(P_1 + P_2)$, which may be expanded by equation (18).

The result of multiplying the two wave-forms, *e. g.* in a suitable modulation circuit, is given by $P_1 \times P_2$. If $\omega_1 = \omega_2 = \omega$, then

$$P_1 P_2 = A_1 A_2 e^{D \cos(\omega t + \alpha)}, \quad (21)$$

where $D = \sqrt{R_1^2 + R_2^2 + 2R_1 R_2 \cos(\phi_1 - \phi_2)} \quad (22)$

and $\alpha = \arctan \left(\frac{R_1 \sin \phi_1 + R_2 \sin \phi_2}{R_1 \cos \phi_1 + R_2 \cos \phi_2} \right). \quad (23)$

Further, the result of sinusoidally amplitude modulating the pulse is simply

$$(1 + m \cos \pi t) A e^{R \cos \omega t}, \quad (24)$$

which obviously contains sidebands of frequencies $(n\omega \pm p)$ about each of the harmonics $n\omega$ of the pulse repetition frequency.

Different types of demodulation processes may be analyzed with the help of equation (21), and the effects of the frequency response of the intermediate transmission path can readily be allowed for, by expansion of the wave-form by equation (18) into its frequency components.

Conclusions.

It has been possible above to consider only the more outstanding analytical features of the exponential pulse. Since pulses of this type can be generated by means of quite conventional circuits, it should be possible to carry out experimental verifications of the theoretical results. Even though the exponential pulse should prove inferior to other pulse wave-forms in practical applications, analysis of the circuit behaviour in terms of the exponential pulse would still appear to possess several advantages.

This paper is published with permission of the Engineer-in-Chief of the Post Office Engineering Department.

Bibliography.

- (1) 'Electronic Engineering' (June, 1942).
- (2) Sullivan, 'Electrical Engineering' (May, 1942).
- (3) James, Polgreen and Warren, J. I. E. E. (August 1939).

XCIV. *The Principles of Carnot and Clausius.*

By Professor W. WILSON, F.R.S.*

[Received July 28, 1943.]

THE following reflections are the outcome of reading a recent paper by Dr. Ehrenberg † and a consequent re-examination of the thermodynamical foundations laid down by Clausius himself. No student of thermodynamics can turn back to the 'Mechanische Wärmetheorie' (especially the later edition, 1876, or its translation, 1879, by W. R. Browne), after reading modern works on the subject, without being greatly impressed and without a renewed feeling of admiration for its author. Like most other great contributors to physical science he owed much to his predecessors and contemporaries—to Carnot, Clapeyron, Joule and Kelvin—and it was no doubt due to their influence, and perhaps partly due to the fashion of his time, that he felt constrained to deduce, or derive, his admirable statement of the second law from a principle, or proposition, that seemed to him more axiomatic or self-evident. This principle, as is well known, he expressed as follows (Browne's translation, p. 78) :—

(A). *Heat cannot, of itself, pass from a colder to a hotter body.*

I have contended, in a recent communication ‡, that it is not a good foundation on which to base the superstructure of thermodynamics, since the words "of itself" need a far too elaborate elucidation. There are also other criticisms which may be made of this principle of Clausius and I am inclined to regard it as a slight (but forgivable) blemish on his great thermodynamical creation. It is useless to attempt to achieve something with it by applying it naïvely, as for example, to the conduction of heat along a bar, and Clausius would appear to have realized this. He makes progress by applying it (see Browne's translation, pp. 79 and 80), in the following revised and more elaborate form (A'), to an old type of reversible process :

(A'). *A passage of heat from a colder to a hotter body cannot take place without compensation.*

From this as a starting point he reaches the conclusion

$$Q_1/Q_2 = \phi(t_1, t_2) = f(t_1)/f(t_2), \quad (1)$$

* Communicated by the Author.

† Dr. W. Ehrenberg : "A Note on Entropy and Reversible Processes," Phil. Mag. xxxiv. p. 396 (1943).

‡ W. Wilson : "Thermodynamical Notes," Phil. Mag. xxxiii. p. 831 (1942.)

in which Q_1 is the quantity of heat communicated to a substance, gas, or water and water vapour, or what not, at the temperature t_1 during reversible expansion from some reversible adiabatic, α , to another, β . Q_2 has a similar meaning. It is associated with expansion between the same adiabatics but at the temperature t_2 . The functions ϕ and f are universal functions of the two temperatures. This important result can be expressed in the following way:—there is a function, S , of the variables of state (*e. g.* volume and temperature) whose differential can be expressed in the form

$$dS = dQ/f(t), \quad (2)$$

and which is of course a *complete* differential, dQ being heat communicated reversibly at the temperature t . The universal function, $f(t)$, depends on the convention adopted to define the temperature scale and is equal to t itself if Kelvin's work scale is adopted. Clausius is not quite clear about the status of $f(t)$, or τ , as he designates it (see Browne's translation, p. 100), since he writes “. . . so that τ is now the *unknown* function of temperature which enters into the Equivalence-value” (my *italics*).

I should have preferred to call (1) or (2) Clausius' principle rather than the statement (A) which Dr. Ehrenberg prefers. Everyone agrees that it is fundamental and unassailable, whereas (A), as it stands, is unusable, and in the form (A') is exposed to criticisms which are set out below.

Equation (1), of course, is Carnot's principle re-stated in a form which shows its compatibility with the energy principle (first law). If Q_1 and Q_2 are both small and if we re-write (1) in the form

$$\frac{Q_1 - Q_2}{Q_1} = \frac{f(t_1) - f(t_2)}{f(t_1)}$$

we get, writing $(\partial p/\partial t)_v dt dv$ for $Q_1 - Q_2$, which, graphically represented on a p, v diagram, is an infinitesimal parallelogram, and dQ for Q_1 ,

$$\frac{(\partial p/\partial t)_v dt dv}{dQ} = \frac{1}{f(t)} \frac{df(t)}{dt} dt,$$

or, if we define C by

$$\frac{1}{C} = \frac{1}{f(t)} \frac{df(t)}{dt},$$

$$\frac{(\partial p/\partial t)_v dt dv}{dQ} = \frac{dt}{C},$$

or
$$dQ = C(\partial p/\partial t)_v dv (3)$$

This equation is of great historical interest and was first given, and rigorously established, by Clapeyron*. The experimental work of Joule and others of his time, combined with this equation of Clapeyron,

* E. Clapeyron : 'Journal de l'Ecole Polytechnique,' vol. xiv. (1834).

constituted the starting point for the great work of Clausius and the Thomson brothers.

Clausius' principle in the form (A) is not, of course, in question. If one end of a copper rod projects into boiling water and the other into a block of ice everybody agrees as to the direction in which heat will flow along the rod. Misgivings arise when we examine the form (A') of the principle and the way in which it is applied. Clausius applies (A') to prove

" . . . that the relation between the quantity of heat carried over, and that converted into work, is independent of the nature of the matter which forms the medium of the change " (see Browne's translation, p. 79).

The proof is well known and is reproduced, with unessential variations, even in quite recent text-books. It has the old feature of one of two reversible engines driving the other backwards*. No particular working substance is assumed to be used in operating them. In the words of Clausius they might be " the perfect gas " in one case and " the combined mass of liquid and vapour " in the other.

Each engine, when working in its normal way, does the same amount of work during a cycle, and when one is driving the other backwards it is helpful to think of the direct and reversed cycles as performed in the same time. It is pointed out that the " two changes, from heat into work, and work into heat, will cancel each other "; and that " in all other respects also the bodies will have returned, at the end of the two operations, to their original condition, with one exception only." The exception referred to is due to the assumption that " if possible " the reversed engine transfers more heat than the driving one, the net result being that heat passes over from the reservoir at the *lower* to that at the *higher* temperature, which of course is in conflict with (A').

But this result shows that in such a process *heat cannot be transferred at all, either from the reservoir at the lower to that at the higher temperature, or in the reverse direction.* Direction does not in fact appear to be involved. After arriving at Clausius' conclusion we can, as explained below, establish that there is a *unique direction* for the transfer of heat; but it is unnecessary to assume this as a premiss. Instead of (A') we may, indeed, use the following statement:

(B). *A passage of heat from one body to another cannot take place by a cyclical process, consisting of a succession of states of equilibrium, which on completion of the cycle leaves no other changes (compensations).*

Having proved (1) by the use of (B) we can easily *infer* a unique direction of transfer of heat when work is done by a reversible engine—since (1) requires that $f(t_1) > f(t_2)$ when $Q_1 > Q_2$.

The axiom preferred by Planck is much superior to (A) or (A'), not only because it is perfectly clear and unambiguous, but because it does

* Clausius does not mention *engines*; but *cyclical processes*.

not contain the unnecessary reference to temperature. It may be stated as follows :

(B'). *It is impossible to construct an engine which performs a cycle of operations periodically, raises a weight, cools a reservoir and does nothing else.*

It can be used in a simple way to establish (1) and to show that there is a unique direction of transfer of heat when an engine, reversible or otherwise, is doing work * *and that this direction is that in which heat travels by conduction.* It contains no reference to reversibility or irreversibility and the impossibility referred to in (B') would become a possibility if the direction of *conduction* of heat were opposite to that of the transfer of heat in the case of an engine doing work.

It is not claimed, be it noted, that Clausius' deduction of (1) and (2) from the principle (A') is unsound. The criticism is, briefly : The statement (A) is useless, while (A') is more elaborate than it need be ; in particular it has an unnecessary reference to the direction of transfer of heat.

The foregoing reflexions lead me to the following observations on Dr. Ehrenberg's thermodynamical arguments. It may be said at once that the condition of reversibility is sufficient for the validity of (2), *i. e.* for $dS=dQ/t$ (assuming Kelvin's work scale and thus replacing $f(t)$ by t). Reversibility is not *always* a necessary condition. What *is* necessary is that the change involving dQ shall consist of a succession of states of equilibrium and we can only regard them as reversible when they are states of *stable* equilibrium. In most cases such a succession of states of equilibrium implies reversibility. The most usual one is that for which

$$dQ=dU+p dv, \quad (4)$$

and hence

$$dS=(dU+p dv)/f(t). \quad (5)$$

Here no mention of any condition is needed, unless indeed it be that the variables of state, t , v and p , shall have a meaning and this implies equilibrium.

In the special case of conduction of heat, or the passage of heat by radiation, without change in volume (4) becomes

$$dQ=dU,$$

and hence

$$dS=dU/f(t). \quad (5 a)$$

Even for this to be valid there must be no more than an infinitesimal variation from thermal equilibrium. Dr. Ehrenberg seems to realize this, for he writes : " Traditionally this can be justified by splitting up the process of heat transfer into a great number of steps, each of which

* W. Wilson, *loc. cit.*

consists of a transfer of heat into or from a reservoir of almost the temperature of the body." But is not this simply the approach to the limiting case of reversibility?

In order to cope with the difficulty that dS is not in fact, in all cases, represented by $dQ/f(t)$ (or dQ/t), Dr. Ehrenberg defines a "body" (for thermodynamical purposes) in such a way as to exclude such cases as the adiabatic expansion of a gas into a previously exhausted space. Obviously the conditions, set out above, for the validity of (4) make this elaborate artifice unnecessary.

One can agree with Dr. Ehrenberg that "the entropy of a single body is simply a function of the state of this body. . . ." This is universally agreed and generally speaking the formulæ which follow from (5), or the equivalent expressions such as

$$dS = (dU + Edq)/f(t)$$

in the case of a cell, formulæ expressing relations between variables of state, *e. g.* between t , v and p , or between t , q and E (temperature, quantity of electricity and electromotive force), or whatever they may be, are not confined to reversible processes.

It is obviously characteristic of the entropy function, S , that the entropy of a closed system (a system in an adiabatic enclosure) cannot change as a consequence of reversible changes or of changes consisting of successions of states of equilibrium, and in consequence we may regard the entropy of such a system as the sum of the entropies of its parts. Dr. Ehrenberg represents them by $\sigma_1, \sigma_2, \dots \sigma_n$. Now Dr. Ehrenberg maintains that Clausius' principle enables us to infer the existence of a single valued function F "so that the question whether a state B can be reached from a state A is answered by the relative values of the function." This, I think, is correct if Clausius' principle has the form I have represented by (A') above. Clausius' principle in the form (A') does more than this: it shows, as we have seen, that, since the entropy of a closed system is the sum of the entropies, $\sigma_1, \sigma_2, \dots \sigma_n$, of its parts, that F must have the form $F(x)$ where $x = \sigma_1 + \sigma_2 + \dots + \sigma_n$, and $\sigma_1 = \int dQ_1/t_1$, $\sigma_2 = \int dQ_2/t_2$ and so on, if $t_1, t_2, \dots t_n$, are Kelvin's work scale temperatures. Now Dr. Ehrenberg *deduces* this result, by an argument which appears to me to be unsound. To begin with he writes

$$F = F(\sigma_1, \sigma_2, \dots \sigma_n),$$

and its form is, so I understand, *not given*, *i. e.* he does not assume in advance that it has the form described above. He then infers, by a process which need not be repeated here, that

$$\frac{\partial F}{\partial \sigma_1} = \frac{\partial F}{\partial \sigma_2} =, \text{etc.} \dots \dots \dots (6)$$

and, *with the aid of this*, that

$$F(\sigma_1, \sigma_2, \dots \sigma_n) = F(\sigma_1 + \sigma_2 + \dots + \sigma_n),$$

where the σ 's have the meanings given above, and also that each σ_v has the form $\int dQ_v/t_v$. Now we can only deduce (6) if the form of F is given in advance. The function F might conceivably be

$$F = \sigma_1 \cdot \sigma_2 \dots \sigma_n,$$

for example, in which case (6) would be untrue. We do not know in advance that it is a function of $\Sigma \sigma_v$, since this is one of the consequences Dr. Ehrenberg sets out to deduce. There appears to me to be a circularity in the deduction. The end result is inferred from (6), which can only be known to be true after the end result has been reached. This kind of thing (I refer only to the *result* of Dr. Ehrenberg's argument) is all contained in the work of Clausius, at least implicitly and in a much simpler and clearer form.

Finally a brief reference to Dr. Ehrenberg's remark "... the second law can be re-stated, that for any process the entropy of a system of two bodies or more increases if it does not remain constant." Here, at any rate, I think, he has the great master Planck on his side*. I have ventured, in the paper already mentioned †, "to lay hands on my father Parmenides" and to criticize this attitude to the second law. For me the second law is Carnot's principle expressed in the form (1) or (2), which appeases the demands of the first law, also a discovery of Carnot ‡, and I regard its directional features as consequential, at all events as they appear in *classical* thermodynamics.

XCV. The Measurement of the Residual Parameters of Variable Air Condensers.

By J. C. SIMMONDS §.

[Received August 25, 1943.]

Introduction.

MAGNIFICATION meters or, as they are more commonly known, "Q" meters, are now in quite general use for power factor measurements on condensers and dielectric specimens at radio frequencies. In many of these instruments the tuning condenser is not of the highest quality and, due to its residual inductance and resistance, serious errors may be introduced. A method of measuring the residual parameters of a variable condenser has been described by D. B. Sinclair ||, and this method can

* M. Planck, 'Thermodynamik,' p. 82 (1905).

† W. Wilson, *loc. cit.*

‡ E. Ariès, L'Œuvre Scientifique de Sadi Carnot, Payot & Cie, Paris (1921)...

§ Communicated by the Author.

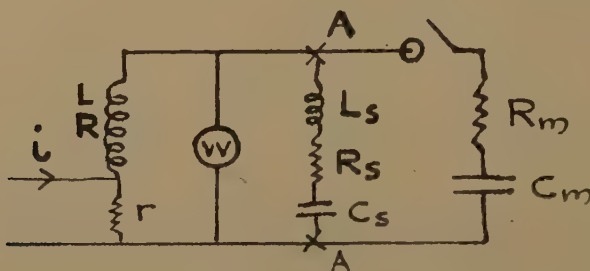
|| Proc. I. R. E. p. 255 (1936).

frequently be used to determine the parameters of the tuning condenser used in a "Q" meter. When this can be done it is, of course, an easy matter to apply corrections to the results of measurements so that greater accuracy is obtained.

An outline of the application of the method to the "Q" meter is given below.

The Principle of the "Q" Meter and the Basic Assumptions of the Method.

The basic circuit of a "Q" meter, as used for a measurement on a capacitance, is shown in the figure. A known current " i " is passed through the small resistance " r ," thus injecting a known e.m.f. into the resonant circuit formed by L and C_s . If L and C_s are adjusted for resonance at the frequency of the current i , a comparatively large voltage appears across the condenser and is indicated by the valve voltmeter "V.V." The ratio of the voltage produced across the condenser



to the voltage injected into the circuit is the magnification of the circuit and, if R_s and r are negligible compared with R , the high frequency resistance of the inductor L , it is also the "Q" of the inductor.

Measurements of the capacitance and power factor of a condenser by means of a "Q" meter are usually made in the following way. The apparatus is tuned to resonance at the desired frequency and the capacitance C_{s1} in the tuning condenser noted, together with the Q, say Q_1 . The unknown condenser of capacitance C_x is then connected across the circuit and resonance restored by reducing the tuning condenser capacitance to C_{s2} . Let the new value of Q be Q_2 . Neglecting the effect of residual series inductance and resistance in the tuning condenser, the capacitance and power factor of the unknown can be found from the equations :

$$C_x = C_{s1} - C_{s2}, \quad \dots \dots \dots (1)$$

$$\text{p.f.} = \frac{C_{s1}}{C_x} \left(\frac{Q_1 - Q_2}{Q_1 Q_2} \right), \quad \dots \dots \dots (2)$$

The residual parameters will make both the capacitance and power factor of the unknown condenser appear too small, as shown below.

The following sections are based upon the assumption that a variable air dielectric condenser can be represented by the equivalent circuit shown between the points AA of the figure, and that the equivalent circuit is not dependent in any way upon the setting of the condenser, except, of course, that C_s is a function of the setting. In other words, it is assumed that the residual parameters can be represented by a series resistance and a series inductance, both being independent of the condenser setting. This assumption is usually justified over a wide range of frequencies. Although L_s is constant over a wide range of frequencies, R_s is proportional to the square root of the frequency.

Determination of the Series Inductance.

The apparatus is tuned to resonance at the desired frequency. Then a suitably small condenser of effective capacitance C_m at the frequency of measurement is connected across the circuit (see the figure), and resonance restored by decreasing C_s . It should be noted that the effect of any residual inductance is included in the value of C_m ; actually, as will be shown below, the value of C_m is not required to be known. Let the suffix "1" refer to values when C_m is not connected across the circuit and the suffix "2" refer to values when C_m is connected across the circuit.

Now the effective capacitance C_e of a condenser, C_s , comprising an actual capacitance in series with an inductance L_s , is given very closely by

$$C_e = C_s(1 + \omega^2 L_s^2 C_s), \quad (3)$$

providing $\omega^2 L_s^2 C_s \ll 1$. Thus, in the two cases just described, the effective capacitances across the points AA of the figure are :

$$C_{e1} = C_{s1}(1 + \omega^2 L_s^2 C_{s1}), \quad (4)$$

$$C_{e2} = C_{s2}(1 + \omega^2 L_s^2 C_{s2}) + C_m, \quad (5)$$

providing $\omega C_s R_s$ and $\omega C_m R_m$ are small compared with unity, which is invariably the case. Since the two effective capacitances are equal :

$$C_{s1} - C_{s2} = C_m - \omega^2 L_s^2 (C_{s1}^2 - C_{s2}^2), \quad (6)$$

C_{s1} and C_{s2} being given, of course, by the low frequency calibration of the condenser.

It appears from this equation that if C_m and ω are kept constant and different values of C_{s1} and C_{s2} used in the measurement, then it will be possible to determine not only L_s but also C_m . Different values of C_{s1} and C_{s2} can easily be obtained by altering the value of L . In effect, therefore, the procedure is to measure a small condenser C_m at different parts of the scale of the tuning condenser C_s , all other quantities except the inductance and resistance of the tuning inductor being maintained constant. When this has been carried out at several different parts of the scale ($C_{s1} - C_{s2}$) can be plotted against ($C_{s1}^2 - C_{s2}^2$). The resulting curve should be a straight line, as from equation (6) it follows that the

relationship between these two quantities is linear, providing the assumptions made are justified. Further, the slope of the line is equal to $-\omega^2 L_s$, whilst its intercept on the "y" axis is C_m .

Once the value of L_s has been found the correct value of any unknown capacitance can be determined by adding $\omega^2 L_s (C_{s1}^2 - C_{s2}^2)$ to the measured value $(C_{s1} - C_{s2})$. It will be observed that no information is required regarding the small condenser C_m in order that L_s may be found and, therefore, this condenser may be allowed to have quite a large residual inductance. The value of L_s is normally independent of frequency over a wide range of frequencies, but if it is desired to apply corrections to measurements made at widely different frequencies, it is advisable to measure L_s at several different frequencies in the range.

Determination of the Series Resistance.

In this section the capacitance values are assumed to be corrected for the effect of the series inductance. The experimental procedure necessary to determine the series resistance of the tuning condenser is almost identical with that of the previous section; the only difference being that the readings of the "Q" meter are noted. In fact, it is quite convenient to combine the two experiments, and to obtain all the information for the determination of both L_s and R_s from one series of measurements.

When the apparatus is tuned to resonance without the condenser C_m connected across the circuit, the total series resistance of the circuit is $R + R_s + r$ and, therefore,

$$R + R_s + r = \frac{1}{\omega C_{s1} Q_1}, \quad \dots \dots \dots (7)$$

since the "Q" of a circuit of the kind under consideration may be defined as the ratio of the reactance of the inductor or condenser at resonance to the total series resistance.

Now, when C_m is connected and the apparatus adjusted for resonance, the total effective series resistance of the circuit is $R + R' + r$, where

$$R' = \frac{C_{s2}^2 R + C_m^2 R_m}{(C_{s2} + C_m)^2}, \quad \dots \dots \dots (8)$$

and the total effective capacitance C' in series with R' is

$$C' = C_{s2} + C_m, \quad \dots \dots \dots (9)$$

providing both $\omega C_{s2} R_s$ and $\omega C_m R_m$ are small compared with unity. Therefore,

$$R + R' + r = \frac{1}{\omega C_{s1} Q_2}, \quad \dots \dots \dots (10)$$

since the total capacitance across the circuit is $(C_{s2} + C_m)$, which must be equal to C_{s1} . From equations (7), (8) and (10) it is seen that

$$\frac{C_1}{C_m} \left(\frac{Q_1 - Q_2}{Q_1 Q_2} \right) = \omega R_m C_m - \omega R_s (C_1 + C_{s2}), \quad \dots \dots \dots (11)$$

where $C_m = C_{s1} - C_{s2}$.

All the quantities involved in this equation are known with the exception of R_m and R_s . If the measurement is now repeated, as in the previous section, at different parts of the tuning condenser scale, $\frac{C_{s1}}{C_m} \left(\frac{Q_1 - Q_2}{Q_1 Q_2} \right)$ can be plotted against $(C_{s1} + C_{s2})$. The curve obtained is a straight line, provided R_s is independent of the setting of C_s , and its slope is $-\omega R_s$. It will be observed that again no information is required regarding the condenser C_m in order that R_s may be found. In fact R_m may be obtained from the graph, for the intercept on the "y" axis is $\omega R_m C_m$. Since the series resistance R_s is a function of frequency it should be determined at several frequencies.

From equation (11) the power factor of an unknown condenser, C_x say, is seen to be given by

$$\text{p.f.} = \frac{C_{s1}}{C_x} \left(\frac{Q_1 - Q_2}{Q_1 Q_2} \right), \quad \dots \dots \dots (12)$$

if R_s is zero. If R_s is not zero, a correction given by $\omega R_s (C_{s1} + C_{s2})$ must be added to the power factor calculated from equation (12).

Conclusions.

The residual inductance and resistance of the tuning condenser of a "Q" meter can be determined from observations made on the meter itself, the only additional equipment required being a fixed condenser of fairly small capacitance. It is not necessary to know the effective capacitance or power factor of the fixed condenser; indeed, they may be found as part of the measurement.

XCVI. Units and Dimensions.

By HAROLD JEFFREYS*.

[Received October 30, 1943.]

THE most gratifying feature of the recent discussion in the 'Philosophical Magazine' and the 'Proceedings of the Physical Society' is that hardly any contributor has anything to say for the assertion, which still appears in textbooks for mathematical students, that the ratio of the electrostatic and electromagnetic units of charge is the velocity of light. On the other hand, nobody has mentioned explicitly the chief reason for rejecting it, that electric charge is what Campbell has called a fundamental

* Communicated by the Author.

magnitude, but now prefers to call an independent magnitude *. It is capable of physical addition. In electrostatics the forces produced by neighbouring charges are additive; electric currents side by side are additive, and the charges passed in different intervals of time by the same current are additive. But the charge on a Leyden jar, measurable by electrostatic methods, can also be used to deflect a ballistic galvanometer, and therefore is also measurable by the methods of current electricity. Theoretically at least we could build up by the proof plane an electric charge on a sphere that would be measurable in the latter way.

The situation is strictly similar to the measurement of distances alternatively in centimetres or in miles. In either case the methods of fundamental measurement will compare the same magnitude with either standard; and if there is any case for giving different dimensions to the e.s. and e.m. units of charge, a precisely similar case could be made out for giving different dimensions to a centimetre and a mile; and then there ceases to be much use in having dimensions at all.

Without prejudice to the possibility that extensions of the definition may be useful, I shall say that quantities comparable with the same quantity by fundamental measurement are *of the same kind* and must have the same dimensions. There is nothing metaphysical about this, as some contributors seem to think; the definition is entirely in terms of possible physical procedures.

The choice of a unit is a convention. The insertion of the usual adjective "mere" is a mistake. There is an arbitrary element in a convention; but when two or more conventions are made they must satisfy rules of consistency; it is also desirable that they should be convenient. Inconsistency can arise with as few as two conventions. We do use the word *red* to denote the colour of a tomato; people with a different language might use it to denote the colour of grass, perfectly consistently; but if it was used in both senses at once it would lead to the inference that a tomato is the same colour as grass, which is false, whatever language is used to express it. In electrical theory three constants have been given numerical measure 1 by choices of units, but the further principle (I do not decide at present whether it is hypothesis or convention) that they are numbers has led to the contradiction that two comparable magnitudes have different dimensions. Consistency, therefore, can impose a limitation on the choice of conventions. It is inconsistent to measure different quantities of the same kind in different units without stating what the units are; consistency is saved here, possibly at the cost of some inconvenience, by stating the units explicitly.

* Phil. Mag. xxxiii. p. 763 (1942). I think that both adjectives are liable to misinterpretation, and would prefer *additive* to either.

On this subject I welcome the remarks of Dr. Guggenheim *, which agree with those that I made in 'Scientific Inference' †. We can say that 10 cm. are the same length as 3.97 inches, thus using what I called the language of quantity in describing what I should call a property of an object. This usage is convenient and consistent, which is as much as we have any right to ask. The property is described by the complete expression 10 cm., not by either part, the 10 or the cm., separately. But it follows that the dimensions of the length are characteristic of the length itself; they persist when the unit changes. It is obviously convenient to use a unit of the same kind as the quantity being measured, but when we speak of dimensions we are speaking of the quantities themselves and only indirectly of the units. An example of a unit not of the same kind as the quantity being measured is the millimetre used to express pressure. (Meteorologists use the millibar, which really is a pressure.) The same facts could certainly also be described in the idealist language that Professor Dingle tries to use, but I am not at all sure that he has said how to do it. Dingle (Phil. Mag. xxxiii. p. 696, (1942)) says explicitly that he does not regard a length as an intrinsic property of the body, which is discovered by means of the standard. That is, apparently, if comparison fails to disclose any difference between a set of bodies by means of the operation chosen, he refuses to abstract the common property of the class and call it the length. This is not a matter of metaphysics but of language; Dingle can presumably dispense with abstract nouns if he wants to, at the cost of intolerable circumlocution.

Professor Dingle ‡ uses "fundamental" in a different sense when he calls a thermometer reading a fundamental measurement. Temperature is not capable of physical addition. A thermometer reading is the ratio of two lengths multiplied by 100, and is a number. The attribution of dimensions to temperature is a further step, but when it is taken the temperature is no longer necessarily identical with the thermometer reading. The question here is, why is it worth while? The only explanation that I can see is that statistical mechanics requires a parameter that appears in various consequences of the theory in the same way as temperature appears in experimental laws. This parameter is an energy per degree of freedom. If it is identified with temperature, the latter is thereby made into an energy and must have the dimensions of the latter. The thermometer reading, being intrinsically a ratio, can equally well be the ratio of two energies and there is no inconsistency. Apart from this theory there seems to be no reason for regarding temperature as anything but a number. If we are satisfied to regard the theory of

* Phil. Mag. xxxiii. pp. 479-96 (1942).

† I still think that chapters 4 and 6 to 9 of this book are worth reading; what is good in the others is done better in my 'Theory of Probability.'

‡ Phil. Mag. xxxiii. pp. 321-44 (1942).

heat as a study of thermometer readings we need go no further. But since they are ratios they might be the ratios of any magnitudes of the same dimensions, and the fact that a thermometer reading is a number does not prove that temperature is. It merely shows that it may be, and leaves the actual dimensions indeterminate until further evidence is available, and in this case the further evidence comes from the fact that the theory works.

The situation with regard to Coulomb's law is similar, up to a point. Denoting the constant factor by κ , we get an unambiguous determination of the dimensions of κq^2 . It remains undecided how these are shared between κ and q^2 . We can, if we like, take a new dimension Q for charge, as Duncanson * does, and put a corresponding Q^{-2} into κ ; or we can follow Maxwell and take κ as a number and get a determination of the dimensions of q †. But, unlike those of temperature, the dimensions of q are not suggested by anything else. The question therefore reduces to one of convenience, and of how far we can follow convenience without landing in inconsistency.

Now one important criterion for convenience is the existence of a reproducible standard to serve as a unit. This is true of length, time, and mass. This consideration affords a strong objection to attempts to use, for instance, the velocity of light and the law of gravitation to determine two of the fundamental units in such a way that they will be germane to the other. (I think that this word of Guggenheim's is an improvement on the usual *consistent* for this purpose.) The measurement of distance by the travel-time of light is a very difficult task. If ordinary measurements had to be done in this way it would take days or weeks to get the accuracy that at present takes seconds. The same applies to the use of such a unit of mass as would produce an acceleration of 1 cm./sec.² in a particle 1 cm. away, and, of course, to the further use of Planck's constant to eliminate the last of the fundamental units. But every difficult comparison of this sort inevitably increases error, and measurements in different comparisons with the ideal standard will not be comparable among themselves to anything like the accuracy that can be attained with the present standards. An alternative would be to compare the present standards with the ideal standards as accurately as possible, to continue to make measurements with the present standards, but to apply constant factors to reduce to ideal standards. But the uncertainty of the factors would still usually be greater than that of the direct measurements. Hence the uncertainties of different results would be strongly correlated, and comparison would be rendered much more difficult than it is at present. The practical advantages of expressing the velocity of light, the constant of gravitation, and Planck's constant in terms of the present standards, rather than conversely, are overwhelming. These considerations have nothing to do with dimensions.

* Proc. Phys. Soc. liii. p. 432 (1941).

† Or we can take Q to have dimensions of force, as Dingle does.

The three universal constants are themselves dimensional, and if their dimensions are taken as basic those of length, time and mass will be determined from them.

The decision about the best electrical standards is a matter for those directly concerned, but the following suggestions are offered for consideration. The coulomb, I think, is better standardized than the electrostatic unit of charge and would probably be the best fundamental unit. The most accurate method of determining magnetic fields seems to be by comparison with those produced by electric currents; in that case there seems to be no reason for maintaining the unit magnetic pole as an independent unit, since moment $\frac{1}{10}\pi$ could be defined as that due to 1 ampere flowing in a circle of 1 cm. radius. This is supported by the theoretical consideration that magnetism in general, in the last analysis, seems to be due to moving electrical charges. This would suggest taking the fundamental law of electricity and magnetism to be that of the force between two electric currents *in vacuo*. The constant factor could be absorbed without contradiction into the dimensions of electric current, leading to a determination of the dimensions of Q . Magnetic moment will have dimensions QL^2/T . The last differs from Duncanson's result by a factor μ (the permeability of a vacuum); but there seem to be advantages in absorbing μ into the magnetic moment anyhow.

This process amounts to treating the fundamental magnitude q as a quasi-derived one, but there is no objection to this provided that one realizes what has happened. The same procedure is followed when angle, which is measurable as a fundamental magnitude in Euclid's way, is interpreted as the ratio arc/radius in consequence of the experimental laws connecting distances. In the latter interpretation angle is a number, and we can without inconsistency regard the measure in degrees as the same number multiplied by $180/\pi$. But the elimination of the apparent dimensions of angle has been made possible only by means of experimental laws and is not simply a matter of definition.

Dr. Campbell has suggested an additional dimension to express shape. Nobody would deny the importance of shape; but does the notion of dimension help to express it? At the worst, physical quantities have dimensions in length, mass and time that are of the forms n or $n + \frac{1}{2}$, n being integral. But if we introduce a dimension for shape and take a sphere as having shape-dimension 0, we shall need a continuous set of two dimensions to cover the range of ellipsoids, and more for more complex forms. Such complexity hardly seems worth while. I think that the language of dimensions is unsuitable for talking about shape, and a suitable language exists, namely that of geometry.

Some contributors have expressed a wish for an extended theory of dimensions that will express vector properties; they ask that a Cartesian co-ordinate should have different dimensions from a length. What does not seem to be realized is that all physically observable quantities are scalars. Directional properties by themselves do not indicate a vector, otherwise the potential due to any non-spherical distribution would be

a vector. To take another illustration from Professor Dingle, in the Doppler effect $\delta\lambda/\lambda$ depends on the direction of motion of the star observed. But vector properties concern transformations from the system of reference used by one observer to that used by another, not changes of the system observed. There is all the difference in the world between passing on one's results to another observer, who will be able to translate them into his own language, and observing a different system oneself. This type of confusion saturates present-day relativity theory. The point is that in this case Professor Dingle's line of sight passes through himself and a given star, and the velocity in his line of sight is the scalar product of the direction cosines of this line and the velocity of the star. Any other observer translating into his own system would also take the scalar product. What he would *not* do is to point his telescope in some other direction and think that he was observing the same star; the preoccupation with vectors implies that he would. Vector properties are important simply on account of the fact that they are useful in expressing the scalar properties of Euclidean mensuration, replacing many quadratic and higher relations between distances by linear ones between co-ordinates. Vector properties, however, are derived from scalar ones, not conversely, and the dimensions of a vector are simply those of its magnitude multiplied by three numerical factors expressing its direction.

XCVII. *On the Continuous Spectrum of Hydrogen-like Atoms*
(parabolical co-ordinates).

By W. R. MORGANS, D.Sc.*

[Received May 24, 1943.]

1. EXPRESSIONS for the intensities in the continuous spectrum of the hydrogen atom for transitions between the continuous (hyperbolic) stationary states into the discrete (elliptic) ones have been determined⁽¹⁾ when the co-ordinates are spherical polar. In the treatment of problems involving capture of electrons, it is of advantage to use parabolical co-ordinates. Expressions for the intensities for transitions between the continuous and the discrete states when the co-ordinates are parabolical co-ordinates, have not, as yet, been determined. The object of the present note is to calculate these intensities for the Lyman and the Balmer states, when the co-ordinates are parabolical.

* Communicated by the Author.

2. The Wave Functions.

If we use the parabolical co-ordinates

$$x = \sqrt{\xi\eta} \cos \phi, \quad y = \sqrt{\xi\eta} \sin \phi, \quad z = \frac{\xi - \eta}{2}.$$

$$0 \leq \xi \leq \infty, \quad 0 \leq \eta \leq \infty, \quad 0 \leq \phi \leq 2\pi,$$

and assume the electron and the nucleus as one system, the wave equation for the electron travelling in the direction of the nuclear charge Ze is given by

$$\frac{\partial}{\partial \xi} \left(\xi \frac{\partial \psi}{\partial \xi} \right) + \frac{\partial}{\partial \eta} \left(\eta \frac{\partial \psi}{\partial \eta} \right) + \frac{1}{4} \left(\frac{1}{\xi} + \frac{1}{\eta} \right) \frac{\partial^2 \psi}{\partial \phi^2} + \frac{2\pi^2 \mu E}{h^2} (\xi + \eta) \psi + \frac{4\pi^2 \mu e^2 Z}{h^2} \psi = 0,$$

where ψ is the amplitude function, μ and e the mass and charge of the electron, E the energy of the atomic system, and h Planck's constant.

The discrete state, which we shall represent by dashes, is characterized by the energy $E' < 0$. If we put

$$-\frac{k'^2}{4} = \frac{2\pi^2 \mu E'}{h^2} \quad \text{and} \quad a = \frac{h^2}{4\pi^2 \mu e^2 Z}$$

the solution found which represents the discrete state is

$$\psi(n_1' n_2' m') = e^{-\frac{k'}{2}(\xi + \eta)} (\xi \eta)^{\frac{|m'|}{2}} F(-n_1', |m'| + 1, k'\xi) F(-n_2', |m'| + 1, k'\eta) \frac{\cos m' \phi}{\sin m' \phi},$$

where
$$F(\alpha, \beta, x) = \sum_{p=0}^{\infty} \frac{\alpha(\alpha+1) \dots (\alpha+p-1)}{\beta(\beta+1) \dots (\beta+p-1)} \frac{x^p}{p!}$$

and $|m'| = 0, 1, \dots, \infty.$

We shall put

$$k'(|m'| + 1 + n_1' + n_2') = k'l' = \frac{1}{a}.$$

The normalization factor for the function $\psi(n_1' n_2' m')$ is given by

$$A^2(n_1' n_2' m') = \frac{2a}{\pi} \frac{(k')^{2|m'|+4}}{\Gamma(|m'|+1)^4} \frac{\Gamma(|m'|+n_1'+1) \Gamma(|m'|+n_2'+1)}{\Gamma(n_1'+1) \Gamma(n_2'+1)}.$$

The continuous state, which we shall represent by unaccented letters, is characterized by $E > 0$. If we put

$$\frac{k^2}{4} = \frac{2\pi^2 \mu E}{h^2},$$

the solution found, which represents the continuous state, is

$$\psi(n_1 n_2 m) = e^{\frac{ik}{2}(\xi + \eta)} (\xi \eta)^{\frac{|m|}{2}} F(-n_1, |m| + 1, -ik\xi) F(-n_2, |m| + 1, -ik\eta) \frac{\cos m \phi}{\sin m \phi},$$

where
$$n_1 = \frac{1}{2} \left(\frac{i}{ka} - |m| - 1 \right) + i\zeta, \quad -\infty \leq \zeta \leq \infty; \quad \zeta \text{ real.}$$

$$n_2 = \frac{1}{2} \left(\frac{i}{ka} - |m| - 1 \right) - i\zeta,$$

We shall put

$$-ik(|m| + 1 + n_1 + n_2) = kl = \frac{1}{a}.$$

The normalization factor for the function $\psi(n_1 n_2 m)$ is given by

$$|A(n_1 n_2 m)|^2 = \frac{e^{\frac{\pi}{ka}} k^{2|m|+2} \left| \Gamma \left(\frac{|m|+1}{2} + i \left(\frac{1}{2ka} + \zeta \right) \right) \right|^2 \left| \Gamma \left(\frac{|m|+1}{2} + i \left(\frac{1}{2ka} - \zeta \right) \right) \right|^2}{4\pi^2 (|m|!)^4}.$$

Also

$$\left| \Gamma \left(\frac{|m|+1}{2} + i \left(\frac{1}{2ka} \pm \zeta \right) \right) \right|^2 = \frac{\pi}{\cosh \pi \left(\frac{1}{2ka} \pm \zeta \right)} \prod_{s=1}^{\mu} \left[\left(s - \frac{1}{2} \right)^2 + \left(\frac{1}{2ka} \pm \zeta \right)^2 \right]$$

if $|m| = 2\mu$ (even),

and

$$\left| \Gamma \left(\frac{|m|+1}{2} + i \left(\frac{1}{2ka} \pm \zeta \right) \right) \right|^2 = \frac{\pi}{\sinh \pi \left(\frac{1}{2ka} \pm \zeta \right)} \prod_{s=1}^{\mu} \left[s^2 + \left(\frac{1}{2ka} \pm \zeta \right)^2 \right]$$

if $|m| = 2\mu + 1$ (odd).

The function $\psi(n_1 n_2 m)$ vanishes at infinity as $\frac{1}{r}$, for if we represent $F(-n_1, |m| + 1, -ik\xi)$ in an integral form, then the asymptotic expansion of

$$M(nm\xi) = e^{\frac{ik\xi}{2}} \xi^{\frac{|m|}{2}} F(-n_1, |m| + 1, -ik\xi)$$

can be shown to be

$$M(nm\xi) \sim \Gamma(|m| + 1) \xi^{\frac{|m|}{2}} \left\{ \frac{(ik\xi)^{-\frac{|m|+1}{2}} e^{\frac{ik\xi}{2}}}{\Gamma \left(\frac{|m|+1}{2} + i \left(\frac{1}{2ka} + \zeta \right) \right)} + \frac{(-ik\xi)^{-\frac{|m|+1}{2}} e^{-\frac{ik\xi}{2}}}{\Gamma \left(\frac{|m|+1}{2} - i \left(\frac{1}{2ka} + \zeta \right) \right)} \right\},$$

which, being the sum of two conjugates, is real. Further simplification gives

$$M(nm\xi) \sim 2 \frac{\Gamma(|m| + 1) e^{-\frac{\pi}{2} \left(\frac{1}{2ka} + \zeta \right)}}{\xi^{\frac{|m|+1}{2}} k^{\frac{|m|+1}{2}}} [S \cos \delta - T \sin \delta],$$

$$\text{where } S+iT = \frac{1}{\Gamma\left(\frac{|m|+1}{2} + i\left(\frac{1}{2ka} + \zeta\right)\right)},$$

$$\delta = \frac{k\xi}{2} + \left(\frac{1}{2ka} + \zeta\right) \log k\xi - \frac{|m|+1}{4} \cdot \pi,$$

$$S^2+T^2 = \frac{1}{\left|\Gamma\left(\frac{|m|+1}{2} + i\left(\frac{1}{2ka} + \zeta\right)\right)\right|^2}.$$

But $\psi(n_1 n_2 m)$ is the product of two such similar functions, so that it tends to zero as $\frac{1}{r}$.

3. The Matrices.

The components of the matrices are given by the expressions

$$x(n_1 n_2 m; n'_1 n'_2 m') = A(n_1 n_2 m) A(n'_1 n'_2 m') \int x \psi(n_1 n_2 m) \psi(n'_1 n'_2 m') d\tau,$$

with similar expressions for y and z , where the integral is taken throughout space.

These matrices have been evaluated ⁽²⁾. For the x, y matrices, values exist only for the transitions $m' = m \pm 1$ and the value is given by the expression

$$x(n_1 n_2 m; n'_1 n'_2 m-1) = \frac{\pi}{4} A(n_1 n_2 m) A(n'_1 n'_2 m-1) m^2 (m-1)!^2$$

$$\times e^{-\pi i(n'_1 + n'_2)} \frac{\nu^{2m+2}}{(-ik)} u^{\bar{n}_1 + n_2 + n'_1 + \bar{n}_2 - 1},$$

$$\left[F\left(-n_1, -n'_1; m, 1 - \frac{1}{u^2}\right) F\left(-n_2, -n'_2; m, -\frac{1}{u^2}\right) \right.$$

$$\left. - u^2 F\left(-n_1-1, -n'_1; m, 1 - \frac{1}{u^2}\right) F\left(-n_2-1, -n'_2; m, 1 - \frac{1}{u^2}\right) \right]$$

except in the case $m=1$ when it has double the above expression. The function $F(\alpha, \beta; \gamma, x)$ is the hypergeometric function and

$$u = \frac{k' + ik}{k' - ik} = \frac{l + il'}{l - il'} = e^{2i \arctan \frac{l'}{l}},$$

$$v = \frac{2}{k' - ik} = \frac{2al'}{l' - il'}.$$

The determination of these matrices for all values of m is simplified by the relations

$$x(n_1 n_2 -m; n'_1 n'_2 -m-1) = x(n_1 n_2 m; n'_1 n'_2 m+1) \quad m \geq 0,$$

$$x(n_1 n_2 -m; n'_1 n'_2 -m+1) = x(n_1 n_2 m; n'_1 n'_2 m-1) \quad m \geq 1,$$

$$x(n_1 n_2 m; n'_1 n'_2 m+1) = x(n'_1 n'_2 m+1; n_1 n_2 m).$$

It is to be noted that in the third expression, the matrix component for the transition $m' \rightarrow m+1$ is obtained from the matrix component for

the transition $m' \rightarrow m-1$ by changing m into $m+1$, n_1 into n'_1 , n_2 into n'_2 , n'_1 into n_1 , n'_2 into n_2 , k' into $-ik$, and $-ik$ into k' , due regard being paid to the consequent change in u .

The z matrix exists only for the transitions $m'=m$ and the value is given by the expression

$$z(n_1 n_2 m, n'_1 n'_2 m) = \frac{\pi}{4} e^{-i\pi(n'_1 + n'_2)} v^{2m+4} u^{n_1 + n_2 + n'_1 + n'_2 - 2} (m!)^2 A(n_1 n_2 m) A(n'_1 n'_2 m) \\ \left\{ [(n'_1 - n'_2)(1+u^2) - (n_1 - n_2)(1-u^2)] F\left(-n_1, -n'_1; m+1, 1-\frac{1}{u^2}\right) \right. \\ \times F\left(-n_2, -n; m+1, 1-\frac{1}{u^2}\right) \\ - 2n'_1 F\left(-n_1, -n'_1+1; m+1, 1-\frac{1}{u^2}\right) F\left(-n_2, -n'_2; m+1, 1-\frac{1}{u^2}\right) \\ \left. + 2n'_2 F\left(-n_1, -n'_1; m+1, 1-\frac{1}{u^2}\right) F\left(-n_2, -n'_2+1; m+1, 1-\frac{1}{u^2}\right) \right\}.$$

4. The Intensities.

The intensity of the radiation emitted is given by the expression

$$I = -\frac{64\pi^4 e^2 \nu^4}{3c^3} \int_{-\infty}^{\infty} \Sigma(|x|^2 + |y|^2 + |z|^2) d\zeta,$$

where the integration is taken over the parameter ζ and the summation is extended over all possible states. ν is the frequency of the radiation emitted.

For particular cases of the total quantum number $l' = |m'| + 1 + n'_1 + n'_2$, the previous expressions for the matrices reduce to much simpler forms and the intensities may be evaluated.

5. The Lyman state for which $l'=1$ gives $m'=n'_1=n'_2=0$ and the matrices $x(n_1 n_2 \pm 1, 000)$, $y(n_1 n_2 \pm 1, 000)$, $z(n_1 n_2 0, 000)$, so that the intensity becomes

$$I_l = \frac{64\pi^4 e^2 \nu^4}{3c^3} \int_{-\infty}^{\infty} [4|x(n_1 n_2 1, 000)|^2 + |z(n_1 n_2 0, 000)|^2] d\zeta.$$

Again,

$$4|x(n_1 n_2 1, 000)|^2 = 2^{11} \pi a^3 \frac{l^8}{(1+l^2)^6} e^{\pi l} e^{-4l \arctan \frac{1}{l}} \frac{\left[\left(\frac{l}{2}\right)^2 - \zeta^2\right]}{\sinh \pi \left(\frac{l}{2} - \zeta\right) \sinh \pi \left(\frac{l}{2} + \zeta\right)},$$

and

$$|z(n_1 n_2 0, 000)|^2 = 2^{11} \pi a^3 \frac{l^8}{(1+l^2)^6} e^{\pi l} e^{-4l \arctan \frac{1}{l}} \frac{\zeta^2}{\cosh \pi \left(\frac{l}{2} - \zeta\right) \cosh \pi \left(\frac{l}{2} + \zeta\right)}.$$

But

$$\int_{-\infty}^{\infty} \frac{\left[\left(\frac{l}{2}\right)^2 - \zeta^2\right] d\zeta}{\sinh \pi \left(\frac{l}{2} - \zeta\right) \sinh \pi \left(\frac{l}{2} + \zeta\right)} = \frac{l(1+l^2)}{3 \sinh \pi l},$$

$$\text{and} \quad \int_{-\infty}^{\infty} \frac{\zeta^2 d\zeta}{\cosh \pi \left(\frac{l}{2} - \zeta \right) \cosh \pi \left(\frac{l}{2} + \zeta \right)} = \frac{(1+l^2)}{6 \sinh \pi l}.$$

It thus follows that

$$\int_{-\infty}^{\infty} \Sigma |x|^2 d\zeta = \int_{-\infty}^{\infty} \Sigma |y|^2 d\zeta = \int_{-\infty}^{\infty} \Sigma |z|^2 d\zeta,$$

and that the total intensity is given by

$$I_l = \frac{64\pi^4 e^2 \nu^4}{3c^3} \cdot 2^{11} \pi a^3 \frac{l^9}{(1+l^2)^5} \frac{e^{-4l \arctan \frac{1}{l}}}{(1-e^{-2\pi l})}.$$

$$\begin{aligned} \text{But} \quad \nu &= \frac{E(l) - E'(l')}{h} \quad (l' = 1), \\ &= \frac{\hbar}{2^3 \pi^2 \mu a^2} \frac{(1+l^2)}{l^2}, \end{aligned}$$

$$\text{so that} \quad I_l = \frac{2^{15} \pi^7 e^{12} \mu Z^5}{3 \hbar^6 c^3} \frac{l}{(1+l^2)} \frac{e^{-4l \arctan \frac{1}{l}}}{(1-e^{-2\pi l})}.$$

The coefficient of absorption per atom α_ν is given by

$$\alpha_\nu = \frac{2^9 \hbar^2}{3c\mu^2 Z} \frac{l^7}{(1+l^2)^4} \frac{e^{-4l \arctan \frac{1}{l}}}{(1-e^{-2\pi l})}.$$

6 The Balmer state for which $l' = |m'| + 1 + n'_1 + n'_2 = 2$ gives the x matrices

$$\begin{aligned} x(n_1 n_2 \pm 1, 010), \quad x(n_1 n_2 \pm 1, 100), \quad x(n_1 n_2 0, 00 \pm 1), \\ x(n_1 n_2 2, 001), \quad x(n_1 n_2 - 2, 00 - 1), \end{aligned}$$

so that we need only consider the values of

$$x(n_1 n_2 1, 010), \quad x(n_1 n_2 1, 100), \quad x(n_1 n_2 2, 001), \quad \text{and} \quad x(n_1 n_2 0, 001).$$

It is not difficult to show that

$$\begin{aligned} |x(n_1 n_2 1, 010)|^2 &= 2^{11} \pi a \frac{k'^6 k^4}{(k'^2 + k^2)^8} e^{\frac{\pi}{ka}} e^{-\frac{4}{ka} \tan^{-1} \frac{k}{k'}} \\ &\times \frac{\left(\frac{1}{2ka} + \zeta \right) \left(\frac{1}{2ka} - \zeta \right)}{\sinh \pi \left(\frac{1}{2ka} + \zeta \right) \sinh \pi \left(\frac{1}{2ka} - \zeta \right)} \left[k'^2 - k^2 - 2kk' \left(\frac{1}{2ka} - \zeta \right) \right]^2. \\ |x(n_1 n_2 1, 100)|^2 &= 2^{11} \pi a \frac{k'^6 k^4}{(k'^2 + k^2)^8} e^{\frac{\pi}{ka}} e^{-\frac{4}{ka} \tan^{-1} \frac{k}{k'}} \\ &\times \frac{\left(\frac{1}{2ka} + \zeta \right) \left(\frac{1}{2ka} - \zeta \right)}{\sinh \pi \left(\frac{1}{2ka} + \zeta \right) \sinh \pi \left(\frac{1}{2ka} - \zeta \right)} \left[k'^2 - k^2 - 2kk' \left(\frac{1}{2ka} + \zeta \right) \right]^2. \end{aligned}$$

$$\begin{aligned}
 |x(n_1 n_2 2, 001)|^2 &= 2^{11} \pi a \frac{k'^8 k^6}{(k'^2 + k^2)^8} e^{\frac{\pi}{ka}} e^{-\frac{4}{ka} \arctan \frac{k}{k'}} \\
 &\quad \times \frac{\left[\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2ka} + \zeta \right)^2 \right] \left[\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2ka} - \zeta \right)^2 \right]}{\cosh \pi \left(\frac{1}{2ka} + \zeta \right) \cosh \pi \left(\frac{1}{2ka} - \zeta \right)}. \\
 |x(n_1 n_2 0, 001)|^2 &= 2^9 \frac{\pi}{a} \frac{k'^6 k^2}{(k'^2 + k^2)^8} e^{\frac{\pi}{ka}} e^{-\frac{4}{ka} \tan^{-1} \frac{k}{k'}} \\
 &\quad \times \frac{\left[k'^2 - k^2 + a k^2 k' + 4 a k^2 k' \left(\zeta + \frac{1}{2ka} \right) \left(\zeta - \frac{1}{2ka} \right) \right]^2}{\cosh \pi \left(\frac{1}{2ka} + \zeta \right) \cosh \pi \left(\frac{1}{2ka} - \zeta \right)}.
 \end{aligned}$$

The z matrices are $z(n_1 n_2 0, 010)$, $z(n_1 n_2 0, 100)$, $z(n_1 n_2 1, 001)$, and $z(n_1 n_2 - 1, 00 - 1)$, where

$$\begin{aligned}
 |z(n_1 n_2 0, 010)|^2 &= 2^9 \pi a \frac{k'^4 k^2}{(k'^2 + k^2)^8} e^{\frac{\pi}{ka}} e^{-\frac{4}{ka} \arctan \frac{k}{k'}} \\
 &\quad \times \frac{\left[2k'^2 k^2 + \frac{k'(k'^2 - k^2)}{a} - 4kk'(k'^2 - k^2)\zeta + 8k^2 k'^2 \zeta \left(\frac{1}{2ka} - \zeta \right) \right]^2}{\cosh \pi \left(\frac{1}{2ka} + \zeta \right) \cosh \pi \left(\frac{1}{2ka} - \zeta \right)}. \\
 |z(n_1 n_2 0, 100)|^2 &= 2^9 \pi a \frac{k'^4 k^2}{(k'^2 + k^2)^8} e^{\frac{\pi}{ka}} e^{-\frac{4}{ka} \arctan \frac{k}{k'}} \\
 &\quad \times \frac{\left[2k'^2 k^2 + \frac{k'(k'^2 - k^2)}{a} + 4kk'(k'^2 - k^2)\zeta - 8k^2 k'^2 \zeta \left(\frac{1}{2ka} + \zeta \right) \right]^2}{\cosh \pi \left(\frac{1}{2ka} + \zeta \right) \cosh \pi \left(\frac{1}{2ka} - \zeta \right)}. \\
 |z(n_1 n_2 1, 001)|^2 &= |z(n_1 n_2 - 1, 00 - 1)|^2 \\
 &= 2^{13} \pi a \frac{k'^8 k^6}{(k'^2 + k^2)^8} e^{-\frac{4}{ka} \arctan \frac{k}{k'}} e^{\frac{\pi}{ka}} \frac{\zeta^2 \left(\frac{1}{2ka} - \zeta \right) \left(\frac{1}{2ka} + \zeta \right)}{\sinh \pi \left(\frac{1}{2ka} + \zeta \right) \sinh \pi \left(\frac{1}{2ka} - \zeta \right)}.
 \end{aligned}$$

The integrations with respect to the separation parameter ζ may now be performed by means of the following formulæ. If we put

$$S_v = \int_{-\infty}^{\infty} \frac{\zeta^{v+1} \left(\zeta - \frac{1}{ka} \right) d\zeta}{[e^{\pi \zeta} - e^{-\pi \zeta}] \left[e^{\pi \left(\zeta - \frac{1}{ka} \right)} - e^{-\pi \left(\zeta - \frac{1}{ka} \right)} \right]},$$

then

$$S_0 = \frac{l(1+l^2)}{12 \sinh \pi l},$$

$$S_1 = \frac{l^2(1+l^2)}{24 \sinh \pi l},$$

$$S_2 = \frac{l(21l^4 + 35l^2 + 10)}{840 \sinh \pi l}.$$

Likewise, if we put

$$C_v = \int_{-\infty}^{\infty} \frac{\zeta^v d\zeta}{[e^{\pi\zeta} + e^{-\pi\zeta}] \left[e^{\pi(\zeta - \frac{1}{ka})} + e^{-\pi(\zeta - \frac{1}{ka})} \right]},$$

then $C_0 = \frac{l}{2 \sinh \pi l},$

$$C_1 = \frac{l^2}{4 \sinh \pi l},$$

$$C_2 = \frac{l(1 + 4l^2)}{24 \sinh \pi l},$$

$$C_3 = \frac{l^2(1 + 2l^2)}{16 \sinh \pi l},$$

$$C_4 = \frac{l(48l^4 + 40l^2 + 5)}{480 \sinh \pi l}.$$

Using these values we find that

$$\begin{aligned} \int_{-\infty}^{\infty} |x(n_2 n_2 1, 010)|^2 d\zeta &= \int_{-\infty}^{\infty} |x(n_1 n_2 1, 100)|^2 d\zeta \\ &= \frac{2^{17} \pi \alpha^3}{3 \cdot 5 \cdot 7} \frac{l^9}{(l^2 + 4)^8} e^{-4l \arctan \frac{2}{l}} \frac{e^{\pi l}}{\sinh \pi l} (63l^6 + 455l^4 + 920l^2 + 560). \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} |x(n_1 n_2 2, 001)|^2 d\zeta \\ &= \frac{2^{17} \pi \alpha^3}{3 \cdot 5} \frac{l^{11}}{(l^2 + 4)^8} e^{-4l \arctan \frac{2}{l}} \frac{e^{\pi l}}{\sinh \pi l} (4l^4 + 20l^2 + 15). \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} |x(n_1 n_2 0, 001)|^2 d\zeta \\ &= \frac{2^{16} \pi \alpha^3}{3 \cdot 5} \frac{l^{11}}{(l^2 + 4)^8} e^{-4l \arctan \frac{2}{l}} \frac{e^{\pi l}}{\sinh \pi l} (7l^4 + 40l^2 + 40), \end{aligned}$$

so that the total contribution to the intensity from the sums of the squares of the x matrices becomes

$$\frac{2^{17} \pi \alpha^3}{3 \cdot 5 \cdot 7} \frac{l^9}{(l^2 + 4)^8} e^{-4l \arctan \frac{2}{l}} \frac{e^{\pi l}}{\sinh \pi l} (413l^6 + 2660l^4 + 4240l^2 + 2240).$$

Similarly,

$$\begin{aligned} \int_{-\infty}^{\infty} |z(n_1 n_2 0, 100)|^2 d\zeta &= \int_{-\infty}^{\infty} |z(n_1 n_2 0, 010)|^2 d\zeta \\ &= \frac{2^{16} \pi \alpha^3}{3 \cdot 5} \frac{l^9}{(l^2 + 4)^8} e^{-4l \arctan \frac{2}{l}} \frac{e^{\pi l}}{\sinh \pi l} (43l^6 + 300l^4 + 560l^2 + 320) \end{aligned}$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} |z(n_1 n_2 1, 001)|^2 d\zeta &= \int_{-\infty}^{\infty} |z(n_1 n_2 -1, 00-1)|^2 d\zeta \\ &= \frac{2^{19} \pi a^3}{3 \cdot 5 \cdot 7} \frac{l^{11}}{(l^2+4)^8} e^{-4l \arctan \frac{2}{l}} \frac{e^{\pi l}}{\sinh \pi l} (7l^4 + 35l^2 + 20), \end{aligned}$$

and the total contribution to the intensity from the sums of the squares of the z matrices becomes

$$\frac{2^{17} \pi a^3}{3 \cdot 5 \cdot 7} \frac{l^9}{(l^2+4)^8} e^{-4l \arctan \frac{2}{l}} \frac{e^{\pi l}}{\sinh \pi l} (413l^6 + 2660l^4 + 4240l^2 + 2240).$$

The total intensity is thus given by

$$I_b = \frac{2^{14} \pi^7 e^{12} Z^5}{3 \cdot 5 \cdot 7 \mu \hbar^6 c^3} \frac{l}{(l^2+4)^4} \frac{e^{-4l \arctan \frac{2}{l}}}{(1-e^{-2\pi l})} (413l^6 + 2660l^4 + 4240l^2 + 2240).$$

References.

- (1) Proc. Nat. Acad. Sc. vol. xvi, p. 5 (1929).
- (2) Phil. Mag. vol. xiii. (1932). *Zeits. f. Phys.* lv. p. 725 (1929).

CORRIGENDA.

Dr. Ludwik Silberstein's paper on the "Solution of the Equation $f'(x) = f(1/x)$,"
Phil. Mag. ser. 7, vol. xxx. (Sept. 1940), p. 186.

Line 2, for $2 \cos \left(\frac{\sqrt{3}}{2} t - \frac{\pi}{3} \right)$ read $2 \sin \left(\frac{\sqrt{3}}{2} t - \frac{\pi}{3} \right)$.

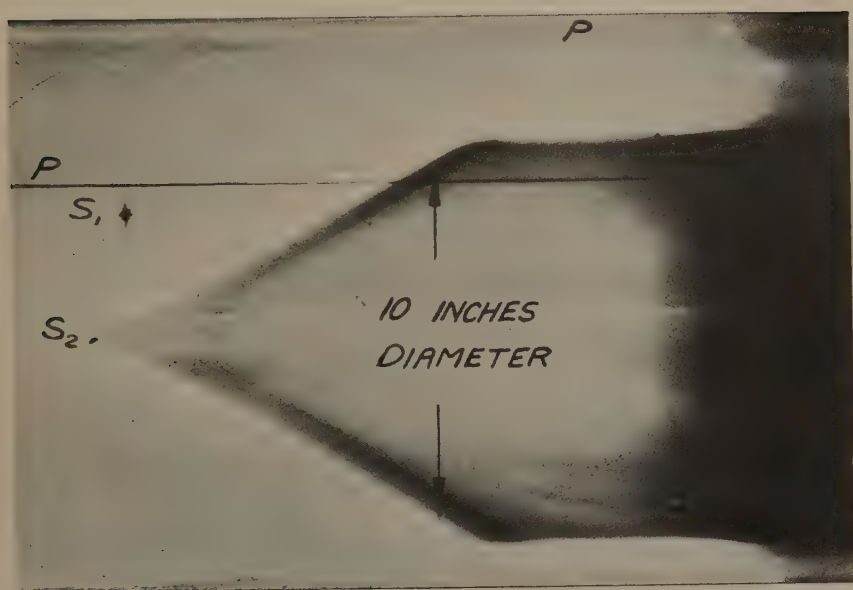
Formula (4), read $f(x) = 2a\sqrt{x} \sin \left(\frac{\sqrt{3}}{2} \log x + \frac{\pi}{3} \right)$,

and, in the following lines, read

$$f(1) = 2a \sin \frac{\pi}{3} = a\sqrt{3},$$

$$f(x) = \frac{2}{\sqrt{3}} f(1) \sqrt{x} \sin \left(\frac{\sqrt{3}}{2} \log x + \frac{\pi}{3} \right).$$

[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]



INDEX TO VOL. XXXIV.

- ADDENDUM. Corrections to Dr. G. F. C. Searle's paper in *Phil. Mag.* vol. xxxiii. Dec. 1942, 360.
- Aharoni (J.), the diode as a rectifier and a detector for weak signals, 505.
- Anomalous viscosity shown in oil flow through engine bearings, 577.
- Archbold (J. W.), preliminary note on condensation in the form of clouds and dew, 632.
- Aristotle, Newton, Einstein, 266.
- Bakerian lecture of 1843, the significance of the, 700.
- Bessel functions of moderate or large argument and order, formulæ relating to, 37.
- Bickley (W. G.), formulæ relating to Bessel functions of moderate or large argument and order, 37.
- Bickley (W. G.) and Miller (J. C.), notes on the reversion of a series, 35.
- Books, new:—The Mathematical Tables Project, 72; Lord Rayleigh, *Life of Sir J. J. Thomson*, 214; Cox, *Photographic Optics*, 360; Dr. J. L. Haughton, *Bibliography of the Literature relating to Constitutional Diagrams of Alloys*, 431; A. J. Thomson, *Table of the Coefficients of Everett's Central Difference Interpolation Formula*, 431; M. Randall and L. Young, *Elementary Physical Chemistry*, 576; R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 576; Prof. F. A. Fish, *Fundamental Principles of Electric and Magnetic Circuits*, 650; Zaki Murri, *Tables of Legendre Associated Functions*, 722; R. A. Archibald, *Mathematical Tables and Aids to Computation*, 794.
- Bradford (S. C.), the properties of fluids, 433.
- Brown (G. Burniston), modern philosophy of science and some problems of classical physics, 26.
- Campbell (N. R.) and Francis (V. J.), voltage fluctuations in resistors (Johnson effect), 259.
- Carnot and Clausius, the principles of, 828.
- Carslaw (H. S.), note on the paper "The temperature distribution around a spherical hole in an infinite conducting medium" by Messrs. Pugh and Harris in the *Phil Mag.* p. 661 (Sept. 1942), 288.
- Chalmers (J. Alan), the separation of electricity in clouds, 63.
- , the single potential difference at a cadmium electrode, 349.
- and Hume-Rothery (W.), contact potentials, 213.
- Clouds, separation of electricity in, 63, 285.
- Complex roots of algebraic equations, the evaluation of, 314.
- Condensation in the form of clouds and dew, preliminary note on, 632.
- Contact angles, 614.
- potentials, 213.
- Continuous beams, free and forced oscillations of: treatment by the admittance method, 49.
- Cornock (A. F.) and Hughes (Joan M.), the evaluation of the complex roots of algebraic equations, 314.
- Corrigenda to Dr. L. Silberstein's paper, vol. xxx. p. 186 (1940), 850.

- David (W. T.) and Mann (J.), unreleased energy in hydrocarbon flame gases, 816.
- Davy (N.), electromagnet polepieces and condensers of special shapes, 803.
- Dimensions of physical magnitudes (Third paper: Electric and magnetic magnitudes), 588.
- Diode as a rectifier and a detector for weak signals, 505.
- Dirac's Representation Theory, difficulties in, 651, 723.
- Displacements produced in Encasté beams due to the elasticity of the supports, approximate theory of, 162.
- Dobbs (H. A. C.), difficulties in Dirac's representation theory, 651, 723.
- Dufton (A. F.), the duple, 607.
- , the warming of walls, 376.
- Duncan (W. J.), free and forced oscillations of continuous beams: treatment by the admittance method, 49.
- Dunstan (A. E.), three types of energy of viscosity, 479.
- Duple, the, 607.
- Ehrenberg (W.), a note on entropy and irreversible processes, 396.
- Electrical pulse, properties of a special type of, 822.
- Electrodynamic volume of suspensions, vapour-pressure and, 101.
- Electromagnet polepieces and condensers with special shapes, 803.
- Electromagnetic waves, the reflexion of, from a parabolic ionized layer, 342.
- Electron density distribution of the upper ionosphere, further notes on, 130.
- Entropy and irreversible processes, note on, 396.
- Erdélyi (A.), inversion formulæ for the Laplace transformation, 533.
- Evans (E. J.) and Powell (Handel), Hall effect and some other physical constants of alloys.—Part VII. The aluminium-silver series of alloys, 145.
- Ferguson (Allan), note on a passage in Stukeley's 'Memoirs of Sir Isaac Newton's Life,' 71.
- Flame temperatures in closed-vessel explosions, photographic determination of, 795.
- Flint (H. T.), the fundamental equation of quantum mechanics, 496.
- Fluids, properties of, 433.
- Fluorescence notation, 212.
- Fluted column, torsion of, 115.
- Francis (V. J.) and Campbell (N. R.), voltage fluctuations in resistors (Johnson effect), 259.
- Freeman (G. F.) on the iterative solution of linear simultaneous equations, 409.
- Friend (J. Newton) and Hargreaves (William D.), viscosities and rheochors of nitric acid, nitro-paraffins and their isomeric nitres, 810.
- , viscosity at the boiling point.—The Rheochor, 643.
- Glasses, visual spectrophotometry of, with special reference to low absorption glasses, 684.
- Graphical differentiation and integration, 565.
- Green (A. E.), a note on stress systems in æolotropic materials.—I. 416; II. 420.
- Green (H. Gwynedd), a note on the motion of a mercury index in a capillary tube, 393.
- Hall effect and some other physical constants of alloys.—Part VII. The aluminium-silver series of alloys, 145.
- Hansel (C. W.), an extension of nomography, 1.
- , graphical differentiation and integration, 565.
- , standard curves, 361.
- Hargreaves (W. D.) and Friend (J. N.), viscosities and rheochors of nitric acid, nitroparaffins and their isomeric nitrites, 810.
- , viscosity at the boiling point.—The Rheochor, 643.
- Heat conduction, steady state problem in, 502.
- Hillman (A. P.) and Salzer (H. E.), roots of $\sin z = z$, 575.
- Hughes (Joan M.) and Cornock (A. F.), the evaluation of the complex roots of algebraic equations, 314.

- Hydrogen-like atoms (parabolical co-ordinates), on the continuous spectrum of, 842.
- Hypogeometric function and the E-function, proofs of some formulæ for, 422.
- Hume-Rothery (W.) and Chalmers (J. Alan), contact potentials, 213.
- Impulse generator, the stage efficiency of. A mathematical note, 353.
- Infinite sets of numbers, on some, 32.
- Infinitesimal bending of a surface, 549.
- Infra-red radiation, the absorption of, by water-vapour and carbon-dioxide, 321.
- Interaction of two point-charges, note on, 712.
- Interferometry, new contributions to. Part I.—New non-localized interference fringes, 555.
- Irons (Eric J.), notes on refraction at a plane interface and by a (triangular) prism, 608.
- , on contact angles, 614.
- Jeffreys (Harold), units and dimensions, 837.
- Keyboard instrument, problem of the, 472; II. 624; III. 674.
- Laplace transformation, inversion formulæ for the, 533.
- Leah (A. Smeeton), the photographic determination of flame temperatures in closed-vessel explosions, 795.
- Linear simultaneous equations, on the iterative solution of, 409.
- Lloyd (Ll. S.), the problem of the key-board instrument, 472; II. 624; III. 674.
- Lowan (Arnold N.), a steady state problem in heat conduction, 502.
- Löwy (H.), on solidified water-films, 67.
- , vapour-pressure and electrodynamic volume of suspensions, 101.
- McCaig (M.), the absorption of infra-red radiation by water-vapour and carbon-dioxide, 321.
- MacRobert (T. M.), proofs of some formulæ for the hypergeometric functions and the E-function, 422.
- Magnetic moments and the vibratory electron, 182.
- Mann (J.) and David (W. T.), unreleased energy in hydrocarbon flame gases, 816.
- Ment (Jack De), fluorescence notation, 212.
- Miller (J. C.) and Bickley (W. G.), notes on the reversion of a series, 35.
- Milne (E. A.), note on the interaction of two point-charges, 712.
- , rational electrodynamics.—I. The limitations of classical electromagnetism, 73.
- , —. II. The ideas of kinematical relativity, 82.
- , —. III. The charge as point singularity, 197.
- , —. IV. The "Radius" of a point-charge, 235.
- , —. V. The neutron and nuclear dynamics, 246.
- Modern philosophy of science and some problems of classical physics, 26.
- Morgans (W. R.) on the capture of a slow-moving directed electron in a Coulomb field of force, 537.
- on the continuous spectrum of hydrogeu-like atoms (parabolical co-ordinates), 842.
- Motion of a mercury index in a capillary tube, note on, 393.
- Murphy (J. W.) and Savic (R.), symmetrical vortex street in sound-sensitive plane jets, 139.
- Neale (S. M.), anomalous viscosity shown in oil flow through engine bearings, 577.
- Newing (S. T.), the approximate theory of the displacements produced in Encastré beams due to the elasticity of the supports, 162.
- Nomography, extension of, 1.
- Nuclear binding energies, note on the analysis of the, 280.
- Nullgeodesics and null-corpuscles in the theory of relativity, 377.
- Peddle (W.), additional note on the trichromatic theory of colour vision (referring to Phil. Mag., Aug. 1942, p. 559), 426.
- , the representation of intensity, hue, saturation, and the magnitude

- of the resultant vector on the trichromatic theory (with reference to recent research on the trichromatic mechanism), 488.
- Pendse (C. G.) on null geodesics and null-corpuscles in the theory of relativity, 377.
- Pincherle (L.), reflexion and transmission by absorbing dielectrics of electro-magnetic waves in hollow tubes, 521.
- Powell (E. O.), an integral related to the radiation integrals, 600.
- Powell (Handel) and Evans (E. J.), Hall effect and some other physical constants of alloys.—Part VII. The aluminium-silver series of alloys, 145.
- Quantum mechanics, fundamental equation of, 496.
- Radiation integrals, an integral related to the, 600.
- Radio Electrodynamics.—I. The limitations of classical electromagnetism, 73.
- . II. The ideas of kinematical relativity, 82.
- . III. The charge as point singularity, 197.
- . IV. The "Radius" of a point charge, 235.
- . V. The neutron and nuclear dynamics, 246.
- Rawcliffe (G. H.), the stage efficiency of an impulse generator. A mathematical note, 353.
- Reflexion and transmission by absorbing dielectrics of electromagnetic waves in hollow tubes, 521.
- Refraction at a plane interface and by a (triangular) prism, notes on, 608.
- Representation of intensity, hue, saturation, and the magnitude of the resultant vector on the trichromatic theory (with reference to recent research on the trichromatic mechanism), 488.
- Repulsive action between a heated solid surface and the adjacent gas.—Part I. 120.
- Resistors, voltage fluctuations in (Johnson effect), 259.
- Reversion of a series, notes on, 35.
- Rheochor, The.—Viscosity at the boiling point, 643.
- Rheochors of nitric acid, nitro-paraffins and their isomeric nitrates, viscosities and, 810.
- Roberts (F. F.) and Simmonds (J. C.), some properties of a special type of electrical pulse, 823.
- Robertson (Ian A.), magnetic moments and the vibratory electron, 182.
- Roots of $\sin z=z$, 575.
- Rydbeck (Olof E. H.), further notes on the electron density distribution of the upper ionosphere, 130.
- , the reflexion of electromagnetic waves from a parabolic ionized layer, 342.
- Salzer (H. E.) and Hillman (A. P.), roots of $\sin z=z$, 575.
- Savic (P.) and Murphy (J. W.), symmetrical vortex street in sound-sensitive plane jets, 139.
- Sensory energy, with special reference to vision and colour-vision, 289.
- Shaxby (John H.) on sensory energy, with special reference to vision and colour-vision, 289.
- Silberstein (Ludwik) on some infinite sets of numbers, 32.
- , infinitesimal bending of a surface, 549.
- , corrigenda to paper in vol. xxx. p. 187 (1940), 850.
- Simmonds (J. C.), the measurement of the residual parameters of variable air condensers, 833.
- and Roberts (F. F.), some properties of a special type of electrical pulse, 823.
- Simpson (Sir George C.), the separation of electricity in clouds, 285.
- Single potential difference at a cadmium electrode, 349.
- Slow-moving directed electron in a Coulomb field of force, on the capture of, 537.
- Solidified water-films, 67.
- Sound-sensitive plane jets, symmetrical vortex street in, 139.
- Spectrophotometric investigations on iron oxide-containing soda - lime - silica glasses.—Part I. The influence of concentration and ferri-ferrous dissociation, 744.

- Spencer-Gregory (H.), the repulsive action between a heated solid surface and the adjacent gas.—Part I. 120.
- Spontaneous fluctuation voltage, distribution in time of, 716.
- Standard curves, 361.
- Stevenson (A. C.), the boundary couples in thin plates, 105.
- , the torsion of a fluted column, 115.
- , some boundary problems of two-dimensional elasticity, 766.
- Stress systems in æolotropic materials.—I. 416; II. 420.
- Stukeley's 'Memoirs of Sir Isaac Newton's Life,' note on a passage in, 71.
- Surdin (M.), distribution in time of spontaneous fluctuation voltage, 716.
- Temperature, pressure, and specific volume changes of a gas under dissociation and reassociation conditions, 486.
- "The temperature distribution around a spherical hole in an infinite conducting medium" by Messrs. Pugh and Harris in the *Phil. Mag.* p. 661 (Sept. 1942), note on the paper, 288.
- Thin plates, boundary couples in, 105.
- Tolansky (S.), new contributions to interferometry.—Part I. New non-localized interference fringes, 555.
- Trichromatic theory of colour vision, additional note on (referring to *Phil. Mag.*, Aug. 1942, p. 559), 426.
- Tucker (W. S.), the determination of velocity of sound by the employment of closed resonators and the hot-wire microphone, 217.
- Turner (W. E. S.) and Wang (T. H.), some spectrophotometric investigations on iron oxide-containing soda-lime-silica glasses.—Part I. The influence of concentration and ferric-ferrous dissociation, 744.
- Two-dimensional elasticity, some boundary problems of, 766.
- Units and dimensions, 837.
- Unreleased energy in hydrocarbon flame gases, 816.
- Vand (Vladimir), a note on the analysis of the nuclear binding energies, 280.
- Variable air condensers, measurement of residual parameters of, 833.
- Velocity of sound by the employment of closed resonators and the hot-wire microphone, determination of, 217.
- Viscosity, three types of energy of, 479.
- Walker (W. J.), temperature, pressure and specific volume changes of a gas under dissociation and reassociation conditions, 486.
- Wang (T. H.), the visual spectrophotometry of glasses, with special reference to low absorption glasses, 684.
- and Turner (W. E. S.), some spectrophotometric investigations on iron oxide-containing soda-lime-silica glasses.—Part I. The influence of concentration and ferric-ferrous dissociation, 744.
- Warming of walls, 376.
- Whittaker (E. T.), Aristotle, Newton, Einstein, 266.
- Wilson (W.), the principles of Carnot and Clausius, 828.
- Winter (H. J. J.), the significance of the Bakerian lecture of 1843, 700.

PRINTED IN GREAT BRITAIN BY TAYLOR AND FRANCIS, LTD..
RED LION COURT, FLEET STREET, LONDON, E.C.4.